

ON THE LIMITS OF EXPERIMENTAL PSYCHOLOGY¹

BY B. A. FARRELL

Wilde Reader in Mental Philosophy, University of Oxford

Wittgenstein has said: 'The existence of experimental methods [in psychology] makes us think we have the means of solving the problems which trouble us; though problem and methods pass one another by askew [windschief aneinander vorbei laufen].'² As far as I can discover, Wittgenstein does not produce anything to support this contentious and obscure remark. I shall use it as my text, and try to indicate where it is useful and where it is quite unhelpful. In doing so I shall take you for a logical walk (or, more probably, a trot) and resume the topic of the interesting but, to me, inconclusive symposium on Experimental Psychology which was held at the Annual General Meeting at Liverpool a few years ago.

TWO FEATURES OF PSYCHOLOGICAL WORK

I begin by reminding you of two features that characterize the results of psychological endeavour.

In the course of the past eighty years or so, psychologists have discovered a great deal about what might loosely be called 'the adaptive behaviour' of organisms. These discoveries are to be found in the thick forests of the literature. What psychologists have not yet discovered is how to synthesize or unify their discoveries. That is to say, they have not succeeded in hitting on some generalizations by means of which, to put it loosely, they will be able to explain the mass of varied and particular discoveries they have made. For the generalizations that express the discoveries they have made are relatively low level, or specific, in character. What they have not yet succeeded in doing is to find sufficiently high-level generalizations, by means of which they can derive, roughly speaking, their specific generalizations and the particular statements embodying their discoveries.

Well, this is a familiar, indeed a commonplace, feature of psychological endeavour. The second and related feature I wish to mention is perhaps not so clearly appreciated, but, since it takes some time to describe it adequately, I shall merely indicate what it is. As we all know, it is characteristic of many psychological generalizations that we can find exceptions to them. It is notorious how vulnerable they are apt to be. The second feature, therefore, I wish to mention is the vulnerability of generalizations in psychology. It is obvious that these two features of psychological work are closely related, but I shall not pause to examine their relationship here.

EXPERIMENTATION AS THE SOLUTION

How, now, can this present situation in psychology be remedied? Psychologists have taken it for granted that this situation *will* be remedied in the long run. In other words, they have presupposed that remedial action is possible *in principle* and that, consequently,

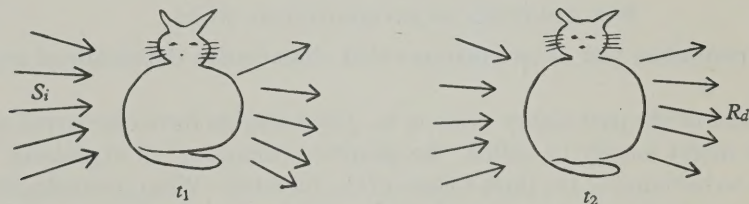
¹ This article is based closely on a paper read at the Annual General Meeting of the British Psychological Society, at Nottingham, April 4th, 1954.

² Wittgenstein (1953), part II, xiv. This unsatisfactory translation is my own.

there is nothing absurd in supposing that we will one day light upon the high-level generalizations which will unify our psychological knowledge, and replace generalizations which are vulnerable by others that are in effect definitive. This presupposition is not obviously true, and I am not even sure that I know what it means. But I shall accept it for the purposes of this paper and pass on.

So the question is: How can this situation be remedied in *practice*? What psychological method, or family of methods, will be sufficient to do so?

When one stands on the high shores of this island and breathes in the transatlantic winds that beat upon them, one gets the impression that some psychologists take the line that what is sufficient to remedy the situation are the standard, or typical, procedures of psychological experimentation, when executed in the usual context of what is commonly called the hypothetico-deductive method. By the typical procedures of psychological experiment, they mean, I think, to refer to something quite orthodox. They mean to refer to a type of experimental design that can be represented in the following sort of text-book diagram (after Andrews, 1948, chap. 1):



Here we consider an organism during any one sufficiently short slice of its life history, t . The organism is thought of as being bombarded, or disturbed, by a large number of stimuli, or stimulus conditions, or what you will. It is also thought of as making various adjustments or responses at this time. Now some special condition of disturbance is introduced at t_1 , those consequences we are interested in. This is labelled the independent variable, S_i . We then set about discovering what response, R_d , or whether some responses R_d , at some later phase of the organism's life history (say t_2 or t_4) is a function of the independent variable or not; or in what way it is a function of it. In the course of doing this we are weighed down with care trying to keep everything else constant, arranging for controls, and so forth. Our overall concern in this sort of experiment is with connexion between disturbing (or stimulus) conditions on the one hand, and behaviour, or conduct or performance on the other.

Of course, this picture gives a misleading impression in various ways of the typical standard procedure of psychological experiment that many psychologists have in mind. For example, it suggests that the stimulus conditions always come from the external environment and that the dependent variables are necessarily overt or publicly observable performances. Still, I hope the picture is good enough for my purposes. What is more important is to notice what the picture excludes. It excludes the use of differential methods. It excludes much of the work done in what is loosely called 'Physiological Psychology'. Thus, much of the work reported in chapters II to VIII in Stevens' *Handbook* (Stevens 1951) would not pass. For work on 'Excitation and conduction in the Neuron', 'Synaptic Mechanisms', 'Sensory Mechanisms' and 'Motor Systems'—such

work does not consist typically, if at all, of discovering how some organismic performance is connected with some certain independent variable. Clearly, too, this picture of the standard psychological experiment excludes the typically clinical investigation, designed to discover, for example, how lesions of the parietal lobes are connected with disturbances of spatial perception, and so forth.

I have said one gets the impression that some psychologists seem to take the line that the typical procedures of psychological experimentation, in the context of the hypothetico-deductive method, will be sufficient to remedy the situation. I won't bore you with supporting quotations (but see Hull, 1943; Skinner, 1953); instead I will go on at once to ask: Is this belief well founded? This question, and all questions of this sort are notoriously difficult to answer for the well-known reason that we cannot tell in advance what scientific strategy will do the trick. If one advocates a certain type of scientific strategy, one is in effect backing a certain horse to win. What, therefore, we are doing in answering this question is to say whether it is reasonable, or sensible, to back the horse of experimental psychology or not. I shall set about answering this general question in the following way. First, I shall ask and try to answer this query. What do high-level generalizations look like where they have been achieved, and where, therefore, our knowledge in some particular field has been unified in some large or considerable measure? Is there any general feature that these high-level statements characteristically exhibit? I shall then ask a second question. Are there any reasons to suppose that the high-level generalizations we aim at in psychology will *not* exhibit this feature also? If there are no such reasons, if, in other words, psychology seems to be analogous to other sciences in this respect, I shall go on to a third question: Is experimental method in psychology, as typically depicted and used, likely to give us the high-level generalizations we require? If it will, then the belief of our transatlantic colleagues will have good support. If it will not, then we will have a good reason for rejecting their belief.

HIGH-LEVEL GENERALIZATIONS IN SCIENCE

I begin with the first query. What do high-level generalizations look like where they have been achieved? Here I shall speak very ignorantly and tentatively, and I shall have to rely a good deal on what others tell me.

The first thing to emphasize, I think, is that the high-level generalizations which have been achieved do not all look alike. Consider, for instance, those which were achieved in Mechanics. These are generalizations about physical things like levers, pulleys, freely falling bodies, the moon and the planets and the like; and they state certain invariances that these macroscopic objects exhibit. This sort of generalization is important for us, all other reasons apart, because some psychologists appear to have been influenced by the example of progress in Mechanics, and to have supposed that the high-level generalizations of psychology will also be of this sort (e.g. Hull, 1943, chap. II). Now, *if* they are going to turn out to be of this sort, then a good case can be made out for saying that experimental method will be sufficient to give us them. But I think a good case can be made for the view that the high-level generalizations of psychology will *not* look like this at all, and that the example of Mechanics is quite misleading. There are three obvious reasons for holding that this example is misleading: (i) There are the wide differences between the phenomena exhibited by the sun and the moon, etc., and the phenomena of adaptive behaviour.

(ii) The generalizations of Mechanics are, perhaps, uncharacteristic of those usually achieved when a science progresses. (iii) The high-level generalizations which have been achieved in fields more closely related than Mechanics to the field of adaptive behaviour are of a different sort to those found in Mechanics. Consequently, when I now go on to say what the high-level generalizations of science look like, I shall pick on and stress a type different from that found in Mechanics. I pick on this different type because it does seem to be more *characteristic* of the generalizations achieved in the course of scientific progress, and because it is the sort of generalization that has been achieved in fields not too far removed from that of adaptive behaviour. I shall introduce this different type in the following way.

When one looks at the history of science, it seems a rather striking fact that progress is often rapidly achieved when the workers in any given field are able to hit upon generalizations which state invariances between minute things (or constituents) or classes of things (or constituents), or state an invariant property of some minute things or classes of things (or constituents). Thus, chemistry is transformed as soon as it hits on the minute things and their properties upon which the phenomena of combustion and chemical combination depend. Biology in general gets under way when it can produce generalizations about microscopic entities, called 'cells', and explain the phenomena of reproduction, growth, metabolism and the like by means of the properties of these, and lesser, microscopic entities. Physics has managed to unify large parts of its field in the same way; and, where it has not yet done so, we are strongly tempted to say that this is because it has not yet hit upon the fundamental particles and/or their properties necessary to achieve the synthesis required. In short, it is a striking and fairly common, but by no means universal, feature of the high-level generalizations in science that they are concerned with the minute constituents of macroscopic objects or things, and with the properties of these minute constituents.

This is all a familiar, indeed a commonplace story. But it is also a vague story as it stands, and it is necessary to make some of it a little more precise. For this purpose I shall use some distinctions related to some suggested by Broad (1933, chap. xiv, § 3).

Consider what would correctly be called 'a thing' or 'an object', or 'a sample of a particular thing or substance'. This thing can be treated as a *simple* or as a *compound* thing. Thus, when I occasionally make the morning porridge, I treat the salt I put into the saucepan as a *simple* thing, in contrast with the porridge which it goes to make up. But when a chemist considers a given quantity of salt, he regards it as a *compound* thing, consisting of two other sorts of things or constituents, namely sodium and chlorine ions related in a cubically structured lattice. To this chemist it is the atoms of sodium and chlorine, with the respective positive and negative charges, that are the *simple* constituents of salt.

Next, a compound thing may have dispositional properties which the simple things composing it do not have. For example, common salt is salty to taste, but sodium and chlorine are not. A piece of sodium at room temperature is solid, but a sodium atom is not. Following Broad, I shall call a property which belongs to a compound thing as a whole, and not to any of its constituents, a *collective* property. A collective property may be of one of two sorts, which I shall explain by a crude and inexact example. Suppose I find some pieces of Meccano lying round the playroom. Suppose I know the dispositional properties of these different pieces, for example, that they are rigid, that they are attachable to each other in various ways, etc. Suppose, next, that I consider putting the pieces

together in a certain special way. They will then form a compound thing, with possibly some collective properties, for example, that the compound thing is able to move smoothly along a level surface carrying small loads—in other words, is able to act like a lorry. Now, it may be possible for me to infer, or deduce, this collective property from my knowledge of the separate pieces of Meccano, and from my knowledge of the special relations which I am supposing these pieces have to one another in order to form this particular compound thing. In other words, consider the generalization ‘All compound things, of which this is an instance, have the collective property, viz. of being able to move smoothly, etc.’. This generalization follows logically from the generalizations I have at my disposal about the separate pieces of Meccano and about the properties of these pieces in the special mutual relations considered. Where this is the case, I shall call the collective property, following Broad, ‘a reducible property’. These contrast with collective properties which cannot in principle be deduced from a knowledge of these constituents in the special relations involved in making up a particular compound thing. Broad calls these ‘emergent properties’. For example, common salt is ‘salty’, but neither of its constituents are; and it can be argued that this property, ‘salty’, is not deducible in principle and therefore is emergent.

One more piece of logical tooling, though I am afraid it is a very rough piece. Consider and contrast a hydrogen atom in a molecule of water, and, say, a leg of a live human being. It is in fact the case that hydrogen atoms can, and do, normally exist in nature in special relations other than those they are in when in water. By contrast, human legs cannot, and do not, normally exist in nature in special relations other than those they are in as part of a human organism. They only have a separate existence in special and infrequent contexts, like those of the anatomy room, the forensic medicine laboratory, and the like. Now, when a hydrogen atom is not standing in the special relations involved in a water molecule, it would be normally natural and true to say of it that ‘it is able to enter into these special relations’, or that ‘it is able to combine with another hydrogen atom and an oxygen atom in a certain way to form a water molecule’. By contrast, since a human leg does not normally occur in nature outside of its special relations in an organism, it is not normally natural and true to say of it that ‘it is able to enter into these special relations’—‘that it is able to combine with other parts of the body to form a human organism’. I shall call the sort of part, of which the hydrogen atom is an example, an ‘independent’ constituent of a compound, in this case water. The sodium of common salt and the pieces of Meccano toy are other examples of independent constituents of their compounds. I shall call ‘a human leg’ a ‘dependent’ constituent of the compound which is the human organism. Many of the constituents of biological compounds appear to be of this latter sort. Of course, this is a rough distinction, as I have already indicated. To speak of what can, and does, normally exist in nature outside of special relations is to speak vaguely; and what it is normally natural and true to say of a constituent of a compound will differ under different circumstances. However, this is all I have the time for, and I think it is good enough for my purposes.

The relevance of the distinction is this. I defined a reducible collective property as one that was deducible from, in part, statements about the properties of constituents when they are *not* in the special relations in which they do stand in the compound thing. From this it follows that, when the constituents concerned are not in these special relations, they must still be able to function with the other constituents of the compound thing to

exhibit the collective property we are deducing. But the only constituents which retain this property when not functioning as part of a compound are the 'independent' constituents of this compound. That is to say, I defined the notion of 'reducible property' by means of the notion of the 'independent constituents' of the compound in question. If, then, one tries to reduce the collective properties of biological compounds, one has to be careful to find independent constituents for the purpose; and this may prove much more difficult and tiresome than has been the case with the collective properties of physical things.

In so far, now, as we have unified our knowledge of nature, we have apparently done so by the use of a few, different sorts of fundamental particles, or minute things, or constituents, and a few properties of these minute constituents. The high-level generalizations we use for this purpose express invariances about these constituents and their properties. By means of them we derive a large part of the remaining generalizations of physics and chemistry, and we thereby explain some of the wide variety of physical phenomena we confront. So when the schoolboy asks us: 'Why is salt soluble in water?', we answer him by deriving this collective property (of solubility in water) from the properties of the constituent ions of salt and molecules of water. We answer him by trying to show how the solubility of salt in water is a reducible property. When he asks: 'But why does it reach a saturation point? And why has the temperature of the water got anything to do with it?', we try to explain the phenomenal variety that salt-in-water can show by reference to the properties of the minute constituents concerned under different conditions.

HIGH-LEVEL GENERALIZATIONS IN PSYCHOLOGY

I turn now with relief from logic and general considerations to the problem about psychology. I have tried to answer the question: 'What do high-level generalizations characteristically look like where they have been achieved?' The next question we set ourselves was this: 'Are there any reasons to suppose that the high-level generalizations we aim at in psychology will be any different in character?'

If they are to be of the *same* sort as those already and characteristically obtained elsewhere, what will they look like? What special form, or forms, will they take? It is clear at once that, if psychologists are to carry out the same sort of programme that has been successfully executed in other fields, they will have to regard 'an organism' in the same sort of way in which it is generally regarded in the biological sciences. They will have to regard it as a compound thing composed of a vast mass of minute things called 'cells' which in turn have other minute things in relations composing each of them. Now the psychologist is specifically concerned with discovering generalizations about the adaptive behaviour, or performances, of organisms. In doing this he is concerned with the various capacities or powers of organisms—the capacity to learn, to remember, to imagine, to see, to solve problems, to tolerate frustration to some extent, to exhibit goal-directed activity, and so on and so forth. It is usual to call these powers or capacities 'dispositions' or 'dispositional properties'. The examples I have listed here are, in an important sense more or less generic dispositions of all organisms or of various species of organisms with which psychologists are concerned. They contrast with the more specific disposition a rat acquires when it has learned to solve a particular maze, or with the one I acquire when I learn the etymology of the word 'Psychology'. In the course of studying 'adaptive

behaviour', psychologists have to handle both the generic and the specific dispositional properties that organisms can and do exhibit. The next point is an obvious one. It is clear that these properties belong to an organism as a compound thing, and not to any of its constituents. Thus, it is not the brain which has the capacity to learn; it is the organism which can do this. It is not any 'cell assembly' which has the specific dispositional property of being able to solve a particular maze; it is an organism that has this property. It is not the occipital lobe which has the power of sight; it is an organism. So in being concerned with certain dispositional properties of organisms, psychologists are concerned with certain *collective* properties which they exhibit.

If, then, psychologists are to unify our knowledge of organismic performance in the same sort of way that has been characteristically achieved elsewhere, they will have to show that the collective properties of organisms with which they are concerned are *reducible* properties. For this purpose they will have to find a few sorts of constituents of organisms and a few fundamental properties they possess—constituents and properties sufficient to enable them to derive the whole range of collective properties that organisms exhibit. Our knowledge about these few constituents and their properties would be contained in a few high-level generalizations expressing the known invariances about these constituents and their properties. From these high-level generalizations we should have to be able, roughly speaking, to deduce, and thereby explain, the wide variety of organismic performance we do confront.

What would a reduction of these collective properties look like? It is only possible at present to give the general form of such a reduction. For we are held up at the outset by not even knowing what candidate, or candidates, to accept for the role of the minute constituents we require. The popular candidate, I suppose, is 'the nerve cell'. Let's adopt it for the present (I shall raise an obvious objection in a moment). Consider the collective property of, say, intelligence, and let me introduce you to two charming white rats, Bertie and Sammie. Suppose we had discovered the properties of some or all nerve cells from which, after investigating the nerve cells of Bertie and Sammie, we could deduce that Bertie was more intelligent than Sammie. We would then be said to have reduced the collective property of 'intelligence'. Or consider some generalization, or tendency statement, about 'intelligence'. For example, 'Rats brought up in good homes, with books, toys and plenty of exposure to intelligent adults tend to be more intelligent than rats brought up in bad homes' (cf. Hebb, 1949, chap. XI). If we knew what the properties of the nerve cells were that determined this difference between the two classes of rats, we should then have available high-level generalizations, which would enable us to explain this fact about differential experience tending to produce differential adult intelligence in rats. Moreover, this difference is more marked with some rats than with others, and we should require our generalizations to explain this fact by reference to differences between the properties of the nerve cells of Bertie, of Sammie and of their relatives.

I can now answer the second question we set ourselves: Are there any reasons to suppose that the high-level generalizations we aim at in psychology will turn out to be of this sort or not? Let us bear in mind that I only ask this question of 'individual' not of 'social' psychology, where quite different sorts of considerations may arise.

There would be good reason for saying: 'No, the high-level generalizations we require will *not* be of this sort', if we had independent evidence for supposing that all the plausible candidates for the role of minute constituents had characteristics that excluded them from

playing this role. For example, if we knew that the nervous system merely acted as an automatic and perfectly regular purveyor of impulses from stimuli to responses, and that nerve cells exhibited no autonomous activity of their own and could not do the job of storage—if we knew that this sort of thing was true of the nerve cells, then these characteristics would probably preclude them from being the minute constituents we want. But, as I understand the position, and my understanding is limited, we have no such knowledge about the nerve cell or any other plausible candidates, if there be any. So I conclude that we have no good reason to suppose that the high-level generalizations we require will *not* take the form I have described. Indeed, it is noteworthy that psychologists who are interested in *how* organisms manage to do the things they do, quite naturally start looking, directly or indirectly, for the parts, or constituents, of their bodies which enable them to perform in this way.

But I think we have *some* reason to suppose that the high-level generalizations we require will not employ 'nerve cells' as the, or one of the, fundamental constituents. For I have already pointed out that the minute constituents or parts we need for reduction have to be independent constituents, and not dependent ones. But 'a nerve cell' appears to be a dependent part, in the circumstances likely to concern us. Consequently, we will not succeed in reducing the collective properties of organisms by means of nerve cells. This suggests that we will not obtain the high-level generalizations we require if we use them. While concentration on the nerve cell will give us something, possibly a great deal, it looks as if all that it *can* give us will be insufficient. It is too gross and dependent an object. However minute a theorist like Hebb is, his fault may be that he is not minute enough (Hebb, 1949). It looks as if we will have to plunge into biochemistry and electrophysiology. I find this a sobering thought. It also looks as if the work of constructing machines which duplicate organismic activity may have a special use in drawing our attention to the *sort* of minute parts of organisms which will be able to play the role of the independent and fundamental constituents we want.

THE INSUFFICIENCY OF PSYCHOLOGICAL EXPERIMENTATION

I shall now apply these remarks to answer the third question I asked: Is experimental method, as typically depicted and used, likely to give us the high-level generalizations we require? If we only use this typical or standard procedure, we seem to be restricted, in principle, and briefly speaking, to two types of generalization. (A) Those connecting stimulus conditions with responses. (B) Those connecting *either* stimulus conditions with some hypothetical internal states of, or occurrences in, the organism, *or* some hypothetical internal states or occurrences with responses, *or* some of these hypothetical states or occurrences with others. I think we have some reason for holding that neither of these two sorts of generalization, A or B, are likely to do the trick.

Consider type A. The words 'stimulus' and 'response' (and their alternates) are admittedly slippery customers, and I am not at all clear how they are used in the literature. As a first move, however, I suggest that they are *not* used to stand for 'constituents' of organisms or other things. The word 'stimulus' is perhaps usually used to refer to some occurrence (or family of occurrences) in the history of other things; or to some feature or set of features exhibited by some thing. This use is good enough to cover the standard cases of buzzers, electric grids, Skinner boxes, tachistoscopic material, and similar

amusements. The word 'response' is usually used perhaps to refer to some occurrence in the life history of an organism, in whatever way this is carved out of its relatively continuous performance. So, however 'stimulus' and 'response' are used in the literature and technical discourse, they do not seem to be used to stand for the minute constituents of organisms, or even for the properties of such constituents. Moreover, since stimuli and responses are indefinite in number and indefinitely variable in character, they will not give us the few fundamental constituents with a few properties which we want. S-R generalizations, therefore, do not appear to be the sort we require.

Consider the second type, B. These generalizations look more promising, since they include something other than stimuli and responses, namely, the internal states of, or occurrences in, the organism; and these states or occurrences may be instances of properties of the sort we want. But our hopes are dashed when we look at some of the internal states and occurrences which have been suggested. Take 'Habit strength'. We are told that this 'exists as an organization, as yet largely unknown, hidden within the complex structure of the nervous system' (Hull, 1943, chap. VIII). This, then, is a property of a large part of the body, namely the nervous system and is far too gross to do the job required. The same applies to internal states or events like that of 'afferent neural interaction'. Alleged occurrences like 'Movement-produced Stimuli' (Guthrie, 1935, *passim*) are rather different. They are manifestations of the properties of certain constituents, namely and in particular, the skeletal muscles, tendons and joints. However, we do not seem to have any evidence to suggest that these are sufficiently fundamental for reductive purposes.

It is worth adding, of course, that there is no reason *in principle* why an experimentalist should not hit upon generalizations which connect, say, stimulus conditions with constituents of the organism which *are* sufficiently fundamental for reduction. There is no reason in principle why he should not postulate the right sort of constituents and properties. The snag is that as long as he sticks to the techniques of experimental psychology, he is unlikely to be in a position to maintain empirical control over statements about these postulated constituents and their properties—a control which Hull rightly insisted he should maintain. He will be unable to test his generalizations, and they will either pass into the sceptical hands of neurologists, or else be relegated to the library of science fiction.

The answer, then, to our third and last question is in the negative, and follows quite naturally from what has gone before. The use of experimental methods alone, as typically depicted and used, will not be sufficient to remedy the present situation and give us the high-level generalizations we want. The confidence of some of our American colleagues in these methods seems to be overdone. At the beginning of this paper I mentioned two related problems that confronted psychologists—the absence of the high-level generalizations needed for unification and the vulnerable character of many of the generalizations they do employ. In so far as psychologists are concerned to deal with these questions, it is true to say that experimental methods, in the typical sense, miss the problems of the psychologist. Problems and methods here do pass one another by askew. In this important respect Wittgenstein seems to have been right. Further, I suspect Wittgenstein may be right in implying that the mere existence of experimental methods makes psychologists think that they have the means of solving the problems which trouble them, including those of unification and vulnerability. But at present I am not clear enough about this implication to pursue it any further.

THE NECESSITY FOR PSYCHOLOGICAL EXPERIMENTATION

Now let me turn briefly to the other side of the picture. Where and how is Wittgenstein's remark unhelpful? The first thing to point out is that the important respect in which Wittgenstein is right is also a very limited respect. Psychologists are in fact concerned with all sorts of *other* problems than the two I have been considering. I suspect that it would be a gross falsehood to say that most of them were actually concerned with this problem at all. It is only ignorant people, like philosophers, who think that psychologists are in active and frantic search for the balm of fundamental law. It is quite unnecessary to draw the attention of a psychological audience to the varied character of psychological problems. Because the typical methods of experimentation may not throw much light on the problems of unification and vulnerability, it does not follow in the least that they throw no light on the *other* problems of contemporary psychology. It does not follow that they pass *these* problems by. Moreover, it is not the case that experimental methods are all of *one* sort. Thus, when I distinguished the typical, or standard, form of such methods, I pointed out that much of the experimental work in chapters II-VIII of Stevens's *Handbook* would not qualify as typical. Wittgenstein's remark suggests that the problems confronting psychologists are all of one sort, and that the experimental methods they use are also all of one sort. This is a very misleading suggestion, and in these respects his remark is quite unhelpful.

But Wittgenstein's dictum is defective in another and more serious way. There is, I have argued, some reason for holding that experimental methods in psychology, in the context of the hypothetical-deductive method, are not *sufficient* to unify our psychological knowledge and give us the invulnerable generalizations we want. But I have not argued that experimental methods are unnecessary for this purpose. On the contrary, I think a good case can be made for saying that experimental methods, while not sufficient, are *necessary*, in some sense, to enable us to unify our psychological knowledge and achieve invulnerability. Wittgenstein's remark is defective, since it makes us overlook this possibility. To say that problems and methods in psychology pass each other by 'windschief' is to suggest that they miss each other *completely*—like two points travelling in different planes—and this is wholly misleading. As far as I can judge, psychologists themselves are apt to be uncertain and unclear about the way in which experimental methods are necessary at this point. Consequently, I shall now try to bring out shortly just why and how they are necessary.

When a chemist claims that the solubility of salt in water has been reduced, he produces a statement like this: 'Sodium and chlorine ions are separable in water.' He uses this statement along with others to explain why salt is soluble in water, and to predict that other substances will or will not be soluble also. I shall call this statement '*p*', and the statement that 'Salt is soluble in water', statement '*q*'. If now a psychologist claims to reduce the 'intelligence' of, say, rats, he would have to produce a statement of the following sort, say: 'Bertie's cortex is capable of carrying a neural load of sort ϕ to the amount F, and this temporal maze imposes a neural load of this sort ϕ to a lesser amount than F.' Let's call this statement '*p'*'. He will have to use *p'* to explain why, or predict that, 'Bertie is intelligent enough to solve this temporal maze'. Let's call this latter statement '*q'*'.

What is the relation between *p* and *q*, and between *p'* and *q'*? It would seem that in

ordinary discourse p and q are not *logically* equivalent to each other, and that p' and q' are not either. Obviously, because when we talk correct English and say that salt is soluble in water, the rules of the use of the word 'soluble' contain no reference to ions, molecules and the like. Similarly, when we say that Bertie is intelligent enough to solve this temporal maze, the rules of use of the word 'intelligent' contain no reference to neural loads of sort ϕ , or to the cortex or to the brain. Well, part of the relation which holds between p and q , and between p' and q' seems to be this. If p is true, then q is also true; and if q is true then p is also true. I shall call this 'equivalence' if anyone wants a name for it, and write it ' $p \equiv q$ '. The same applies to p' and q' . Thus, our neurological reductionist, in search of the cerebral basis of intelligence (as this is usually put), requires his p' and q' to function in such a way that: if it is true that 'Bertie's cortex is capable of carrying a load ϕ to amount F etc.,' then it is in fact also true that 'Bertie has the intelligence to solve this temporal maze'; and conversely.

Now for the point of all this. It seems clear that for a psychologist to *reach* p' , it is not *logically* necessary that he should observe, experimentally or otherwise, the behaviour of Bertie, or any other rat, in a temporal or any other maze. As far as I can tell, the ionic theory of 'solubility' was not reached by making further observations on salt and other substances dissolving in water. It was reached—as far as I can gather—by inference from the theory of crystal structure and the electro-magnetic theory of matter. Similarly, it is quite conceivable that a psychologist could infer from a certain general theory about the cortex and the load of information imposed by environmental situations of different types that rats will be able to solve this temporal maze. But in order to reduce 'intelligence', he has to *show* that $p' \equiv q'$. In other words, to put it roughly and briefly, he has to verify this general hypothesis that if any rat has such and such a cortical structure, then it is able to solve a temporal maze, or, as we would ordinarily and correctly say, that it has the intelligence to solve it. Conversely, he has to show that if a rat has the intelligence to solve this maze, then its cortical structure is of such and such a sort. But it is difficult to conceive of him showing all this without using the typical procedures of psychological observation and experiment. So it seems to be logically necessary that he use them for the purposes of verification.

But more than this. Before a psychologist or neurologist, or any one else, can *reasonably* claim that p' represents a reduction of 'intelligence', he has to have evidence that the manifestations of intelligence, which he claims he can explain by reduction, do correspond roughly with what we normally regard as being manifestations of intelligence. But he can only have this evidence if he, or others, have already carefully explored the manifestations of the collective property of 'intelligence'. How can we reasonably claim to be able to reduce the intelligence of rats (and so find its cerebral basis) if we do not know what rats are capable of doing? And how can we discover this at present except by psychological observation and experiment on rats? Such work is a necessary exploratory step in a programme of reduction. (Just the same, incidentally, is true of 'solubility'. Here the necessary exploratory work on the behaviour of salt in water was done before the theory of crystal structure, etc., was developed; and we are apt to overlook that the same stage was necessary with 'solubility' as is necessary at present with the collective properties of organisms.) Further, though it is possible in principle for a neurologist to light upon a p' without looking at the performances of rats or humans or any other organisms, I think it is a little unlikely that it will be reached like this. At present we are all

very ignorant about the general functioning of the central nervous system. The work of psychologists, experimental and otherwise, is full of suggestions about the way in which the central nervous system must function if organisms are to do the things they can and do do. I think that reductionists will need to make grateful use of all these suggestions, and thereby help to have their attention drawn to the fundamental constituents and properties they require. So experimental work in psychology will also be *practically* necessary in order for us to reach our *p'* statements, and thereby reduce the collective properties of intelligence, vision, learning and so forth.

I hope I have done something to bring out why and how experimental methods are necessary in our search for unity and invulnerability. And I hope I have done enough to show just how defective Wittgenstein's dictum really is.

CONCLUDING SUGGESTIONS

Some concluding points. First, I want to draw attention to some implications of my remarks. *If* psychology achieves its high-level generalizations in the way so characteristic in the history of science, then it seems to follow that much theorizing in psychology over the past thirty years has been misplaced.¹ Much of it has assumed that experimental psychology is strong enough to arrive at Laws of Behaviour—laws which will sit on a base of, at present, unknown neurological generalizations, but which will be synthesized with the latter in due course as these become known. If the analogy holds between the development of psychology and that of the other sciences, then this assumption may be a mistaken one. There may be no laws of behaviour sitting upon a neurological base. The high-level and invulnerable generalizations of psychology may turn out to be quite different in character. It follows, further, that, however great or small the defects in the systems of a man like Hull, no improvements will ever enable them to do the big, grandiose trick we would like them to do.

What sorts of psychological work will probably make the most relevant contributions to the particular problem of high-level generalizations? I tentatively suggest the following: (1) Work in what is loosely called physiological psychology. This extends from work of the Eccles-Adrian sort to that of the Lashley type, and beyond. (2) Clinical work of both the diagnostic and surgical types. If, for example, one investigates the deteriorated performance of people with cortical lesions, one can make suggestions about the way in which parts of cortex and the brain contribute to the normal performance. When one does this, it is not only the case, as Zangwill has rightly said (at Liverpool), that one studies performance 'with the internal environment simplified' in various ways. It is also the case that one is keeping one's feet on the hopeful ground of things and their properties. And it may be much more important to do this than to make a fetish of conforming to the criteria of rigorous experimentation. (3) There is the psychological work which is designed and conducted to throw light on the general plan, or plans, according to which the minute constituents of our bodies mutually co-operate to do their job. I am thinking of work like that by Deutsch (1954). For reasons I have already mentioned there is no need whatever for psychologists to be apologetic about this work, and to grovel in the presence of neuro-anatomists, physiologists and others of their kind.

¹ See, for example, some of the papers reprinted in Marx (1951).

When I compare psychologists with workers in other branches of science, and with my philosophical colleagues, I am forced to say that psychologists in general appear to suffer from methodological hypochondria, and from some personal insecurity about their intellectual standing and professional status in the community. I suspect that *one* of the sources of their hypochondria and insecurity is to be found in the fact that they have not yet managed to make inroads into the two problems of unity and vulnerability that I have picked out for discussion. If we are to accept the suggestions of psychotherapy, it is generally advisable for us to face up realistically to our worries and limitations. I hope that the brisk walk on which I have taken you this morning may do something to help us to face up to one source of our personal uneasiness, and something to show us what may be involved in removing it.

REFERENCES

- ANDREWS, T. G. (ed.) (1948). *Methods of Psychology*. New York: Wiley; London: Chapman and Hall.
 BROAD, C. D. (1933). *Examination of McTaggart's Philosophy*, vol. I. Cambridge University Press.
 DEUTSCH, J. A. (1954). A machine with insight. *Quart. J. Exp. Psychol.* vi, 6-11.
 GUTHRIE, E. R. (1935). *The Psychology of Learning*. New York and London: Harper.
 HEBB, D. O. (1949). *Organization of Behaviour*. New York: Wiley; London: Chapman and Hall.
 HULL, C. L. (1943). *Principles of Behaviour*. New York and London: Appleton.
 MARK, M. H. (ed.) (1951). *Psychological Theory*. New York: Macmillan.
 SKINNER, B. F. (1953). *Science and Human Behaviour*. New York: Macmillan.
 STEVENS, S. S. (ed.) (1951). *Handbook of Experimental Psychology*. New York: Wiley; London: Chapman and Hall.
 WITTGENSTEIN, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

(Manuscript received 30 August 1954)

THE IDEA OF INDEPENDENCE

BY JOHN COHEN AND C. E. M. HANSEL

Department of Psychology, University of Manchester

I. *Introduction* (p. 178). II. *Prediction of a binary event on the basis of previous outcomes* (pp. 178-181). III. *Effect of age on the prediction of outcomes of binary events* (pp. 182-183). IV. *Prediction of a binary event presented visually* (pp. 183-184). V. *Prediction of successive binary events in a series with knowledge of results after each choice* (pp. 184-185). VI. *Factors influencing the prediction of a binary independent event* (pp. 185-187). VII. *Conclusions and discussion of results* (pp. 187-189). VIII. *Appendix* (pp. 189-190). *Reference* (p. 190).

I. INTRODUCTION

We have been intrigued by the fact that the idea of an independent event is a very difficult one to grasp. Many people feel almost compelled to believe that after a run of successes failure is more likely to occur, or the other way round. If, for example, a coin has repeatedly turned up head, the chance of its continuing to do so is believed to get less with each further toss. Even those familiar with notions of statistical independence experience this difficulty. People seem to find it hard to consider an event as separate and detached from a series of similar events in which it occurs, although it may be perfectly evident that the outcome of the event in question can be in no way affected by the outcomes of the other events. The series of events is treated as one continuous single event. The great unlikelihood of getting, say, a run of 10 heads is attached to the possibility of getting a head on the tenth toss, after a run of 9 heads. Similarly, the extreme unlikelihood of not getting a single head in a sequence of 10 tosses is attached to the tenth toss after a run of 9 tails. This sort of reasoning is applied to all kinds of comparable situations in everyday life.

There are countless events in everyday life that can happen in one of two ways. For example, it may rain or not rain on a given day. A child born may be male or female. A candidate may pass or fail an examination. A paper submitted to the editor of a psychological journal may be accepted or rejected. Such events often occur in a sequence. There might be a succession of similar outcomes (e.g. a series of male births) or varying outcomes. In the present study of subjective probability we shall explore the way children at different ages predict the outcome of an event of this character when they are given some information about previous outcomes.

II. PREDICTION OF A BINARY EVENT ON THE BASIS OF PREVIOUS OUTCOMES

In this experiment we are concerned with the prediction of an event that can happen in one of two ways: a bead drawn from a given beaker can be either blue or yellow. The subjects are first given different sequences of previous outcomes and on the basis of these they make their predictions. The sequences are varied in order to discover the effect of particular sequences on the prediction.

The experimenter says to the subjects:

I have here 12 beakers in a row. Each beaker contains blue and yellow beads. The proportions of blue and yellow beads vary from beaker to beaker. I am going to take out 4 beads, one at a time, from each beaker. I will show you the first three beads as they are drawn and then you will guess the colour of the fourth.

The colour of the fourth bead was not revealed.

The experiment was designed such that in any arrangement of 3 beads drawn from a beaker there were always 2 of one colour and 1 of the other in the six possible arrangements. The order of drawing the beads from the beakers was such that there were 2 beads of one and 1 bead of the other colour in the six possible arrangements. Each arrangement was repeated twice, the whole series being given in random order, as below:

Y B B	B B Y	B B Y
Y B Y	B Y B	Y B B
B Y B	Y Y B	Y Y B
B Y Y	B Y Y	Y B Y

The beads were drawn from each beaker in the order reading left to right.

We shall call the colour represented by the two beads the *preponderant* colour, and the colour represented by the single bead the *non-preponderant* colour. The terms preponderant and non-preponderant will be abbreviated as P and $\bar{P}$.

There are three different kinds of arrangement, disregarding colour. The preponderant and non-preponderant events are symbolized as 1, 0 respectively:

Y B B	B Y B	B B Y
or 0 1 1	or 1 0 1	or 1 1 0
B Y Y	Y B Y	Y Y B

Each of the three different kinds of arrangement may have either blue or yellow as the preponderant colour. By combining the arrangements in the manner shown, we remove any possible effect of colour preference. This method of combination enables us to assemble four choices made by each subject for each kind of arrangement. There were 93 subjects yielding 372 choices in all for each kind of arrangement (0 1 1; 1 0 1; 1 1 0). The number and proportion of choices for the preponderant and non-preponderant colour are shown, for each kind of arrangement, in Table 1.

Table 1. *Numbers and proportions of choices of preponderant and non-preponderant colours*

($N=93$. Age 10+)

	$\bar{P}$ followed by P 0 1 1		P in first and third positions 1 0 1		P followed by $\bar{P}$ 1 1 0		Total	
	Y B B or B Y Y		B Y B or Y B Y		B B Y or Y Y B			
Preponderant colour (P)	104	0.27	112	0.30	146	0.39	362	0.32
Non-preponderant colour ($\bar{P}$)	268	0.73	260	0.70	226	0.61	754	0.68
Total	372	1.00	372	1.00	372	1.00	1116	1.00

It may be said that the values shown in Table 1 have been obtained from twelve successive trials with different beakers and that later guesses are affected by earlier trials. We find, however, that a value of 0.27 (for the non-preponderant colour) is obtained in the very first trial of 1 1 0.

The first point to note is that the non-preponderant colour is in each case favoured in the fourth choice. This fourth choice seems to be made in accordance with a principle of com-

pensation or balance which equalizes the possibilities of the two outcomes. It is not made independently of previous occurrences, otherwise there would be no difference in the number of choices of preponderant and non-preponderant colour. Nor is it in accord with a statistical prediction based on sampling from a large population. Two explanations suggest themselves. First, the choice of $\bar{P}$ might presuppose an implicit assumption that an initial store of blue and yellow beads of equal numbers in the beakers has been unevenly depleted. If, for example, a child implicitly assumed that there were at the start 3 blue and 3 yellow beads and he then saw 2 blue and 1 yellow drawn from the beaker, he would presumably conclude that 1 blue and 2 yellow remain and therefore he would prefer to predict yellow. Secondly, and this does not exclude the first explanation, the tendency to favour $\bar{P}$ as the fourth choice may be due primarily to the asymmetrical character of the arrangement of beads already drawn. We have indeed found in previous experiments (Cohen & Hansel, 1955) a tendency to prefer symmetrical arrangements of 4 beads under conditions when there is no question of depleting a store.

Another point to note is that the order of appearance of the colours affects the fourth choice. The more remote the appearance of the non-preponderant colour the more frequently it is chosen. This effect is rather like the one mentioned above as frequently noticed in everyday life in which a person's expectation of heads increases with more appearances of tails.

From earlier experiments (Cohen & Hansel, 1955) we know that subjects tend to assume equal proportions when the proportions are stated to be unknown. If they made such an assumption here, it would not be contradicted by what they saw emerging from the beaker, for if three beads are drawn, a 2:1 ratio of colours is the nearest to a 1:1 ratio that is possible to occur. The fact that they choose $\bar{P}$ suggests that they make this assumption and their choice of $\bar{P}$ enables them to maintain it.

We now turn to experiments in which the results of previous events were communicated to the subjects verbally, and at the same time illustrated visually on a blackboard. The kind of binary event now ranges from independent occurrences at one end to dependent occurrences at the other. The events about which a prediction has to be made are such as come within the experience of children. They include independent occurrences, such as coin tosses and the sex of the next baby to be born in a given town, and dependent occurrences, such as the state of the weather on a given day and which one of two uncles has sent a present. These instances of dependence were selected because it was thought the outcomes would readily be predicted on the basis of previous experience.

The procedure was to say to the subjects:

A boy has two uncles, Uncle George and Uncle John. In 1951 he received a present from Uncle John but not from Uncle George. In 1952, he received a present from Uncle George but not from Uncle John. In 1953 he also received a present from Uncle George but not from Uncle John. In 1954 he has received a present and on the label there are the words 'from Uncle'. Which uncle do you think sent the present?

The appropriate wording was modified to suit the other cases.

The proportions in Table 2 are based on samples of 44. In the two cases marked by an asterisk, $N=88$. All the proportions shown in the table are significant at the 0.01 level. The choice of the non-preponderant event in these experiments is more favoured than in the experiments with beads. This may be due to the tendency to complete or maintain

the pattern that appears in the visual illustration of previous occurrences. For example,

the problem $\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}$ is presented on the blackboard as follows:

	Boy	Girl
1951	✓	—
1952	—	✓
1953	✓	—
1954?		

Moreover, visual illustration also strengthens the tendency to alternate, as may be seen by comparing the middle row of Table 2 with the first and third rows.

Table 2. *Proportions of choices between alternatives in 1954*

Year	P	$\bar{P}$	Winning a toss		Sex of baby		Rain or fine		Sender of present		Mean values	
			P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$
1951	0	1	0.15	0.85	0.12	0.88	0.12	0.88	0.25	0.75	0.16	0.84
1952	1	0										
1953	1	0										
1951	1	0	0.07	0.93	0.10	0.90	0.07	0.93	0.07	0.93*	0.07	0.93
1952	0	1										
1953	1	0										
1951	1	0	0.12	0.88*	0.32	0.68	0.20	0.80	0.23	0.77	0.22	0.78
1952	1	0										
1953	0	1										

* Means of the proportions made when choices are presented in two orders, e.g. $\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}$ and $\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}$, given to two groups of 44 children.

The desire to continue the visual pattern so as to keep or make it balanced may enhance the tendency to choose the non-preponderant value as compared with the choices in the previous experiment with beads. This suggestion is supported by the results obtained when a similar type of problem—in this instance relating to true events—was given without illustration by visual pattern. The extent to which the non-preponderant was favoured in these circumstances was of the order found in the experiments with beads. The problem

was of the form $\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}$ and presented as follows to 89 ten year-old children:

Oxford won the boat race this year and last year. Cambridge won two years ago. Who do you think will win next year?

The choices were 0.38 in favour of Oxford, 0.62 in favour of Cambridge, i.e. the non-preponderant event ($\chi^2 = 4.9$, D.F. = 1, $p < 0.05$).

The alternation or choice of the non-preponderant event also appears significantly, under similar conditions, in the cases $\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}$ (0.36, 0.64) and $\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}$ (0.41, 0.59), but in less

marked fashion. In the cases $\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}$ there appears to be no general tendency to alternate or to favour the non-preponderant event.

III. EFFECT OF AGE ON THE PREDICTION OF OUTCOMES OF BINARY EVENTS

We now consider in relation to one kind of independent event, coin tossing, the predictions made at three ages, 10, 12 and 14 years. These are shown in Table 3. A small number of subjects did not choose P or $\bar{P}$ but replied 'either'. These have been divided between the two choices.

Table 3. *Proportions of choices of alternatives made at different ages*

(10 + ($N=50$), 12 + ($N=64$), 14 + ($N=57$): winning a toss)

Year	P	$\bar{P}$	10 +		12 +		14 +	
			P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$
1951	0	1	0.15	0.85	0.19	0.81	0.29	0.71
1952	1	0						
1953	1	0						
1951	1	0	0.07	0.93	0.23	0.77	0.38	0.62
1952	0	1						
1953	1	0						
1951	1	0	0.12	0.88	0.34	0.66	0.38	0.62
1952	1	0						
1953	0	1						

Of the nine sets of values in this table, seven deviate significantly from equal proportions at least at the 0.05 level of significance. Two values, 0.38, 0.62 (at the age 14 +) are barely significant. They become significant however, if we exclude subjects giving judgements of 'either'. There were 9 in each case, the number remaining out of 57 is therefore 48, of which 17 chose P and 31 chose $\bar{P}$ ($\chi^2=4.3$, D.F.=1, $p<0.05$). It follows that in the circumstances of this experiment there is a decided tendency in the members of the group to predict what has happened less often in the past three events as more likely to happen in the future.

It is clear from this table that between the years 10 to 15 the tendency to favour the non-preponderant event weakens. In the three cases combined there are 150 choices made by the 10 + group, 192 by the 12 + group and 171 by the 14 + group. The numbers of P and $\bar{P}$ choices given by the three groups respectively are:

	10 +	12 +	14 +
P	17	49	60
$\bar{P}$	133	143	111
Total	150	192	171

We can compare the proportions in the 10 + and 12 + groups and in the 12 + and 14 + groups giving the different choices. $\chi^2=10.9$, D.F.=1, $p<0.001$ in the first instance and 3.9, D.F.=1, $p<0.05$ in the second instance. The difference between the 10 + and 14 + is *a fortiori* significant.

This decline with age in the preference for the non-preponderant value in the fourth choice has its parallel in the increased tendency with age to give an 'either' judgement, a tendency which amounts to a recognition of a principle of independence. The ten-year olds did not give a single 'either' judgement, the twelve-year-olds gave 15 (out of 192 choices) and the fourteen-year-olds gave 26 (out of 171 choices) in reply to the three questions.

In all three age groups the tendency to alternate or to choose $\bar{P}$ appears in the cases
 $\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}$ and $\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}$ but not in $\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}$ and $\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}$. Similar results were obtained with the other kinds of
event shown in Table 2.

IV. PREDICTION OF A BINARY EVENT PRESENTED VISUALLY¹

In order to extend the range of conditions under which forms of subjective probability may be investigated and also to standardize a procedure suitable for subjects at all ages we constructed an apparatus consisting of a display board with 16 pairs of lights, as shown in Fig. 1.

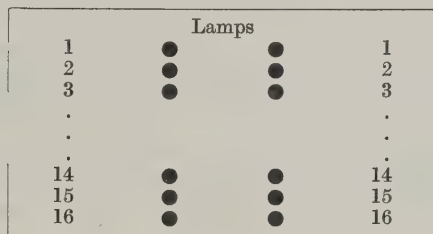


Fig. 1. Display board.

In the experiment, one of each of the first 10 pairs was illuminated, 9 being on one side and 1 on the other. The position of the single light varied (in random order) from position 1 to position 10 in different trials. The preponderant side bears 9 lights and the non-preponderant side a single light. In each set of 10 predictions, made by each subject, the preponderant side was on the right on five occasions and on five occasions on the left.

After seeing the lights 1 to 10 illuminated in rapid succession and remaining lit, the subject had to guess on which side the eleventh light would appear. In fact the eleventh light was never illuminated. The subjects were studied individually. We explained to them that only one of each pair of lights would go on. They indicated their choice by pointing to the small lamp in the eleventh row which they expected to be illuminated next.

The subjects in this experiment were 16 six-year-old and 8 seven-year-old children. Their predictions are shown in Table 4.

Considering first the possible effect of the position in the series of the single light, we did not find, what might have been expected, that the tendency to predict the eleventh light on $\bar{P}$ increased as its position occurred earlier in the series of 10 lights. This may be due to the fact that the display, when all 10 lights were lit, resembled a stable visual pattern and that it was in terms of visual patterns of lights that predictions were made and not in terms of remoteness of occurrence in the past.

Taking next the question whether the non-preponderant side is favoured at all in this experiment, at first sight it does not appear so to any marked extent in the six-year-olds. We found, however, that both the six-year-old and the seven-year-old children tended to alternate their choices of right and left throughout the 10 successive predictions. There were 147 alternating as against 69 non-alternating predictions. In other words, their

¹ We are grateful to Mr R. Church for designing and constructing the display board and for operating it during the experiments.

choices were to a large extent determined by what they had chosen in the previous situation, not by the relations of P and $\bar{P}$ confronting them. On reclassifying the choices into alternating and non-alternating respectively (as in Table 4), we find that when they do not alternate, they do tend to favour the non-preponderant side.

Table 4. *Prediction of eleventh light on preponderant or non-preponderant side*

($N=16$ (6+), 8 (7+).)

Age	Alternation			Non-alternation			Total
	P	$\bar{P}$	Total	P	$\bar{P}$	Total	
6+	58	50	108	13	23	36	144
7+	21	18	39	12	21	33	72
Total	79	68	147	25	44	69	216

On the null hypothesis we should expect an equal number for P and $\bar{P}$ in the non-alternating choices. In fact the number favouring $\bar{P}$ in the combined group of 24 children is significantly large ($\chi^2=5.3$, D.F.=1, $p<0.03$). The tendency to alternate, i.e. to be influenced by choice in a preceding situation, seems to be considerably weaker in the seven-year-old children. These seem more capable of making a choice on the basis of the display confronting them. In so far as they do not alternate the behaviour of the two age groups in relation to P and $\bar{P}$ is much the same.

V. PREDICTION OF SUCCESSIVE BINARY EVENTS IN A SERIES WITH KNOWLEDGE OF RESULTS AFTER EACH CHOICE

The problem here is to identify the influences affecting the prediction of a binary outcome in a series. Among these influences are: (i) the immediately preceding event, i.e. the actual side on which the previous light appeared; (ii) the number of lights which have already appeared on each side; (iii) the immediately preceding choice, i.e. the side *predicted* for the previous light; and (iv) success or failure in the previous prediction.

The apparatus consisted of the display board with 16 pairs of lights already described. This time before any one of a pair of lights was switched on, the subject had to guess which member of the pair it would be. Then one of the first pair of lights was switched on according to a pre-arranged scheme and remained on. The same procedure was followed with each pair of lights in succession from the second to the sixteenth pair. In one form of the experiment the position of the lights was randomized in a 1:1 ratio so that there finally appeared 8 lights on each side in random order; and in the other form of the experiment the position of the lights was randomized in a 3:1 ratio so that finally 4 lights appeared on one side of the display board and 12 lights on the other. There were two groups of subjects, 46 aged 10+ and 16 aged 6-8 years.

From the successive choices in the experiment with a 1:1 ratio in which 8 lights eventually appeared in different orders on each side of the display board the four influences can be assessed.

(i) *Effect of the side on which the previous light appeared*

In both age groups there is a significant tendency to alternate from the side of the previous light. If the previous light is on the right, the choice tends to be on the left and vice versa. χ^2 for the 10+ group is 8.0 (D.F.=1, $p<0.01$) and for the 6-8 group, 20.2 (D.F.=1, $p<0.001$).

(ii) *Effect of number of lights which have already appeared on each side*

Here again at both ages there is a marked tendency to predict the next light on the side with fewer lights. This is a preference for $\bar{P}$. χ^2 for the older group is 32.9 (D.F. = 2, $p < 0.001$), and for the younger group 7.8 (D.F. = 2, $p < 0.05$). From the proportions making the different choices at both ages it is evident that the older children are much more affected by the total number of lights that have already appeared than the younger children.

(iii) *Effect of side chosen for previous light*

In neither of the two age groups is the choice of any side influenced by the side previously chosen. This means that the children are not affected by what they have previously predicted but by the previous outcome.

(iv) *Effect of success or failure in the preceding choice*

Here there is a marked age difference. The younger children alternate after success but not after failure. $\chi^2 = 21.1$ (D.F. = 1, $p < 0.001$). The older children show no tendency to alternate merely because they have either succeeded or failed in their preceding guess.

In the second form of this experiment in which the final ratio of lights was 3:1 we were interested to see whether, when the disparity between the number of lights on either side increased, the children tended to guess the next light as due to appear on the side with more lights. In the early predictions the choices are not unlike those made when the ratio is 1:1. With the 3:1 display, however, a tendency gradually appears to prefer the preponderant side. In the last four choices with the 1:1 display there is no significant tendency to prefer one or the other side, whereas with the 3:1 display there is a significant tendency to prefer P ($\chi^2 = 8.5$, D.F. = 1, $p < 0.01$).

VI. FACTORS INFLUENCING THE PREDICTION OF A BINARY INDEPENDENT EVENT

The following experiment is an attempt to investigate the everyday phenomenon mentioned in the first paragraph that one outcome of a binary event is felt to be more likely as the run of the other outcome is prolonged. The subjects here are girls aged 12–15 years. The two age groups 12+ ($N = 64$) and 14+ ($N = 57$), each comprised two subgroups of higher and lower ability, as judged by the school authorities.

The kind of activity giving the independent event, i.e. coin-tossing, selected for this experiment, has the advantage of resembling a familiar situation. Furthermore, it can be thought of in two ways: (i) in terms of runs of head or tail, (ii) in terms of runs of success or failure of the persons tossing the coin. Two forms of question were used:

1. A girl is tossing a penny. The coin has turned up head on all of the last x tosses. What do you think will turn up the next time it is tossed?

In an alternative form of the above question, 'tail' was substituted for 'head'.

2. Joan and Mary, two captains of school teams, are about to toss a coin before a match. Joan has won and Mary has lost on the last x guesses. Who do you think is more likely to win next time?

In an alternative form of this question the names Joan and Mary were interchanged. In the space marked x a different value from 1 to 10 was placed, a single value being included in the question presented to each subject. Each subject thus made one guess in reply to one of the four types of question. After recording their guess, the subjects were asked to explain why they had given that particular answer.

When we classify the guesses according to length of run of one outcome, we find that 54 out of 121 favour a change in victor or outcome. This number is not, however, significantly different from the number (45) favouring a continuance of the run. Nor does the length of run of one outcome seem to have had any consistent effect on the guess. The proportions favouring each of the two possible outcomes are shown in Table 5.

Table 5. *Proportions favouring continuance or change of run*

Length of run	Continuance of run	Change of outcome	Total
1	0.55	0.45	1.00
2	0.40	0.60	1.00
3	0.45	0.55	1.00
4	0.33	0.67	1.00
5	0.44	0.56	1.00
6	0.55	0.45	1.00
7	0.33	0.67	1.00
8	0.50	0.50	1.00
9	0.43	0.57	1.00
10	0.63	0.37	1.00

In both age groups the proportion of abler subjects giving an 'either' judgement is significantly greater than the proportion choosing one or other outcome. In this respect the abler younger ones resembled the abler older ones more than they resembled their contemporaries.

The fact that we did not obtain a significantly clear-cut tendency in this experiment may be due to the conditions of the experiment: the subjects were not participants in a real situation, they merely answered questions about a hypothetical situation. Other techniques (which we intend to employ) may reveal the alleged effect more definitely.

Our present results suggest that the phenomenon is interwoven with other effects which tend to mask it. The reasons given by the subject for their choices enable us to understand the way this happens. At least 48 subjects stated that they believed it was time for the non-successful outcome to occur, or words to that effect. But in many instances they also adduced other reasons, such as that the tosser was skilful and so the run of successes would continue. Indeed there are a large variety of reasons given and these are often of a highly composite character. It is clear that the decision or choice is far from being a simple one in the majority of instances, and varies considerably from subject to subject. The explanations have been grouped together under the headings: Dependence (including Favouring change, Luck, Determinate, Series and Magic) and Independence.

Some subjects gave more than one reason, so the total number of reasons shown in Table 6 exceeds the size of the sample.

The main features that emerge from the classification of reasons are (i) the relatively high proportion of the younger age-group, as compared with the older, giving 'luck' as the reason for their choice, $\chi^2=4.1$, D.F.=1, $p<0.05$, and (ii) the relatively small proportion of the younger group who say that the outcome cannot be known, $\chi^2=6.8$, D.F.=1 $p<0.01$.

Perhaps the simplest interpretation of these results is in terms of treatment of the unknown outcome as dependent or independent of previous outcomes. The proportions of these two types in the total explanations are about 85 and 15% in the younger group and 73 and 27% in the older group. Those giving explanations in terms of dependence offer two antithetical elements. Some children treat the dependence so as to lead to continuance of the same outcome. Other children treat the dependence so as to lead to a change of outcome. The former group use the data as providing information about the

Table 6. *Frequency of reasons given for prediction of outcome*

Dependence	Age	
	12-13*	14-15†
Favouring change	24	24
Luck	23	11
Determinate	16	18
Series	6	8
Magic	3	4
Total	72	65
Independence	13	24
Total	85	89

* $N = 32 + 32 = 64$.

† $N = 29 + 28 = 57$.

personal characteristics of the tosser such as skill, determination or luck or about the characteristics of the coin, and they extrapolate this to the next throw. The latter group consider the unknown outcome as determined by previous outcomes but not because these previous outcomes reveal the characteristics of the tosser or the nature of the coin. The mere fact that an outcome has occurred once or more times or that one person has had a run of luck means for them that there will be a change of outcome. In the simplest form this is expressed in the notion of alternation, e.g. 'I believe Mary will win next time because it goes alternately'.

VII. CONCLUSIONS AND DISCUSSION OF RESULTS

The conclusions of the various experiments described above are briefly summarized:

(i) (a) After seeing 3 beads each of which may be one of two colours drawn from a beaker containing unknown proportions of blue and yellow beads, the non-preponderant colour is favoured in the fourth choice.

(b) The order of appearance of the colours affects the fourth choice, the more remote the appearance of $\bar{P}$ the more frequently it is chosen.

(ii) In comparable experiments with varying kinds of event the previous outcomes of which are visually illustrated, $\bar{P}$ is even more strongly favoured than in the foregoing experiments with beads.

(iii) $\bar{P}$ is more favoured in independent kinds of event such as coin-tossing and less favoured (as compared with coin-tossing) in events where the outcome appears strongly dependent on previous outcomes (e.g. 'sender of present').

(iv) Between the ages of 10 and 15 years the tendency to favour $\bar{P}$ weakens and the tendency to give 'either' judgements increases.

(v) (a) In experiments with children aged 6+ and 7+, using a display board with 9 lights on one side and the tenth light occupying any position from 1 to 10, we found a marked tendency to alternate choices from right to left in the six-year-olds but scarcely any such tendency in the seven-year-olds.

(b) In so far as they do not alternate they tend to favour $\bar{P}$.

(vi) In experiments with 6+, 7+ and 10+, using the display with a 1:1 ratio of lights appearing at random, we found the following tendencies:

(a) To change from the side of the previous light.

(b) To change from the side with more lights to the side with fewer lights.

(c) An age difference in the effect of success or failure in the preceding choice. The 6-8 group alternate after success but not after failure. The older children show no tendency to alternate merely because they have either succeeded or failed.

(vii) In experiments similar to those in (vi) above, but employing a 3:1 ratio of lights, the early predictions resembled those obtained under 1:1 conditions. The later predictions tended in the direction of the side with more lights.

(viii) (a) In experiments with groups of children 12+ and 14+ on prediction of an unknown binary outcome after varying length of run of one outcome, no striking increase in the tendency to change was found as the length of run was prolonged. This was interpreted, on the basis of subjects' protocols, as due to two antithetical ideas of dependence in this situation. Sometimes the antithesis was present in the explanation of one and the same person.

(b) Fourteen-year-olds gave more 'either' judgements than the twelve-year-olds, and the abler children at both age levels gave more 'either' judgements than the relatively less able.

(c) The frequency of explanations of choice in terms of 'luck' declined and in terms of indeterminateness increased in the 14+ as compared with the 12+ group.

We may now attempt to trace the development of the idea of independence in so far as it emerges in these experiments. At the youngest age which we have investigated, 6 to 7 years, the most marked feature is a tendency to alternate from the previous choice in situations of binary outcome. To the extent that they do not alternate their choices they are apt to choose the non-preponderant outcome. From their remarks it seems that they regard this as fair. They are also influenced by whether they have succeeded or failed in the last attempt: after success they alternate, after failure they do not. To a relatively slight extent they also tend to alternate from the side of the previous outcome and from the side with more outcomes. In short the six-year-olds show a powerful tendency to alternate as a result of a variety of specific influences.

Children a year older are much less apt to alternate their choice. Like some of the six-year-olds but to a more marked degree, the seven-year-olds tend to change from the side of the previous outcome and from the side with more outcomes to the side with less outcomes. But unlike the younger ones they do not tend to alternate from previous choice and mere success or failure in the preceding choice does not influence them. They are much more under the influence of the display confronting them.

We have found that most children, at least till the age of 10+, do not initially distinguish between an event the outcome of which is in reality independent of previous outcomes, such as tossing a coin, and an event with a varying degree of dependence on previous outcomes, such as fine or not-fine weather in a given place on a certain day.

Judgements on these two types of situation are similar for short runs of, say, eight events. $\bar{P}$ is favoured even when knowledge of results shows that one event arises three times as often as the other. Children may treat such short series of independent and dependent events without distinction as single unit events, hence favouring $\bar{P}$. This effect is very marked in series consisting of five or six events.

In longer series, however, they do seem to make the distinction. When the event is of the coin-tossing (i.e. independent) kind, the children seem to predict in longer series in much the same way as they do in shorter series. When, however, as with the display board, the event in the early part of the series may be equally regarded as either of one kind or of the other, the predictions in the later part of the series (after approximately eight outcomes) follow the preponderant outcome. This happens when the ratio of outcomes is 3:1 and would no doubt be even more marked with higher ratios.¹

An idea of independence manifested in a judgement of 'either' does not appear in our experiments at the age 10+. It first emerges at the age 12+. From then onwards the tendency to favour the non-preponderant outcome weakens, together with a strengthening of the inclination to give 'either' judgements.

Decisive changes in the idea of independence as held by some subjects take place between the ages of 12+ and 15+. The idea of luck has a variety of forms and meanings at this period. Gradually it appears to merge in the notion of chance as a concept of independence. For instance a child says: 'It is just luck.' There are two opposing ideas of independence at this time sometimes held by the same subject. One of these ideas of independence favours continuance of a series of similar outcomes. The other favours a change to the other binary outcome as the series proceeds. The idea of independence is beginning to emerge at the age 12+ and is clearly and forcefully held by some subjects at the age of 14+. Even at this age, however, the vast majority entertain ideas of dependence; the idea of luck is still potent, and ideas of magic are not lacking.

VIII. APPENDIX

Some typical explanations of the reasons for the choice of the outcome in section VI are given below:

Favouring change

'Because if Mary has won 7 times she is not likely to win again.'

'Mary (will win) because Joan won all the other times and it is about time Mary won.'

'I think Mary will win because on the Law of Averages it is Mary's turn to win.'

'Joan will probably win the toss because usually the one who lost the time before will win the next time.'

'Because usually the one who has been losing a lot wins. But really it is just luck.'

Luck

'One girl has won 4 times but her luck has run out so the second girl wins and the first girl loses because tails has not won.' (This also favours change.)

'Mary seems to be the lucky one of the two and she won first time.'

'Joan seems to have all the luck so I'm sure she will win again.'

'I think it would be tails because most people think it is the luckier of the two.'

¹ If all the lights were on one side (i.e. in the ratio 1:0) we should have a simple learning situation.

Determinate

'Because Mary most likely can only throw a certain way and Joan knows what it usually comes to

'Mary won the last two so she must be the best and if she is the best she will surely win the next one

'Because if Joan has won 6 already she is obviously better than Mary and so it is possible that she will know more the next time.'

'Because the wind might be causing the coin to drop on the side that it has been doing previously

'Joan (will win). Joan would be very determined to win again and if the same person tossed she would probably call what she called the last time which was heads. Mary would be very nervous and have given up hope of winning.'

Series

'Tails. Because it has turned up tails 3 times already and it is sure to turn up the same after three times turning up tails.'

'I think heads will turn up next time because it has for 5 times.'

'It will turn up Heads because if it has turned up the same ten times before it will probably do so again.'

'Mary is likely to win because the coin usually turns up six or more times the same, if it has turned up five.'

Magic

'Mary will think she will win and so probably will not win.'

'Heads. Because for some strange reason, the fifth toss always comes heads. I read this in a book, and tried it often, and it always works (book title, *1000 Wonderful Things*).'

'If the girl chose tails I think it will be tails because there is a saying that "Tails, tails never fails". This saying is often very true.'

'Tails. Because 7 is my lucky number and it is the seventh time it has been tossed after 6 heads. So I think it will change to tails by the 7th time.' (This also involves the idea of luck.)

Indeterminate

'It depends on the way the coin is tossed up. Any one of them might win. Mary has won the guesses but this does not mean that she will win this time. She might very well but Joan could win this time. It is not possible to tell who will win.'

'Heads or tails. It's just up to the penny.'

'Because if a penny spins in the air it does not necessarily land on the same side every time.'

'Because a penny has two sides, a head and a tail, and no one can tell which end it is going to land (unless they cheat) so it could be either heads or tails.'

'Because when you toss a penny nobody, *nobody* can tell even the cleverest of people which it is going to be, because it might be either.'

'Either. It just depends on which side it lands after being tossed.'

REFERENCE

COHEN, J. & HANSEL, C. E. M. (1955). The idea of a distribution. *Brit. J. Psychol.* XLVI, 111-21.

(Manuscript received 1 July 1954)

PROBLEM-SOLVING AS DEPENDENT ON AVAILABILITY OF FUNCTIONS¹

By PER SAUGSTAD

University of Oslo

I. *Introduction* (pp. 191–193). II. *Experimental* (pp. 193–196): (i) *Subjects* (p. 193); (ii) *The problem* (pp. 193–194); (iii) *Procedure* (pp. 194–195); (iv) *Results* (pp. 195–196). III. *Discussion* (pp. 196–197). IV. *Summary and conclusion* (pp. 197–198). *References* (p. 198).

I. INTRODUCTION

There exists, as yet, no satisfactory definition of what constitutes a 'problem'. Perhaps the nearest approach to such a definition is that of Humphrey (1951): 'A problem is a situation which for some reason appreciably holds up an organism in its efforts to reach a goal' (p. 312). However, whatever the nature of a problem, it is fairly certain that its solution involves the use of past experience and the combination of certain 'functions'—or to use Harlow's (1951) phrase, 'organized response patterns'—which, as part of the problem situation, are initially discrete. The purpose of this paper is to explore the possibility of analysing problem-solving in terms of 'functions', with particular emphasis on their availability to the subject at the time of attempting the solution and on their combination.

The present study arose out of a consideration of the work of Maier (1930, 1933, 1937, 1945) and is perhaps best introduced by a consideration of Maier's pendulum problem. The task was to construct two pendulums, each of which would swing over a designated point on the floor and make a chalk mark at that point. Various materials were available to the subjects: a large table, two long poles, two shorter poles, one table clamp, two burette clamps, two lengths of flexible wire, eight pieces of lead tubing and several pieces of chalk. The solution involved three steps: (1) constructing the pendulums by inserting a piece of chalk into a clamp tied to the end of a piece of wire, (2) making a pole long enough to reach the ceiling by clamping two poles together, and (3) supporting the two pendulums at either end of a third pole which was then wedged against the ceiling with the two joined poles. To ensure that his experimental group had the necessary experience to solve the problem (i.e. the necessary knowledge of the three steps involved), Maier had previously shown them (1) how to make a plumb line out of a burette clamp and a pencil, (2) how to combine two shorter poles into a longer one by clamping them together, and (3) how to wedge a piece of cloth against the wall of a door-way by pressing the side of a flat board against it and then keeping this in place by wedging another board horizontally between the flat board and the other wall of the door-way. The subjects in the group given this pre-problem experience did not solve the pendulum problem although, argued Maier, they had sufficient knowledge to do so. Weaver & Madden's (1949) repetition of this experiment gave essentially the same results.

Now, in asserting that the above subjects had all the experience necessary to the solution of the problem, Maier was rather rash. After all, a particular experience always

¹ This study was conducted while the writer was Rockefeller fellow in Psychology at the University of Chicago. The article is based upon a dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Chicago.

involves a specific, concrete situation, and it is from a succession of such particular experiences that principles, concepts, universals (whatever we choose to call them) are derived. Maier demonstrated the function of certain objects in one definite situation, but to solve the problem, other objects had to be used in another situation. The pre-problem demonstrations would only assist the subjects in solving the pendulum problem if they were seen as illustrating principles of fairly wide applicability. Thus, the situation in which the cloth is wedged against the door-post by the flat board is not the same as that in which a pole is wedged against the ceiling. Even if the sight of the pole elicited the principle of wedging there still remained the problem of finding a second flat surface (the ceiling). The reality of this remaining problem is obvious from the performance of a second experimental group to which Maier gave so-called 'direction' by demonstrating how the pendulum problem might be solved by nailing the wire to the ceiling. More members of this group solved the pendulum problem than of the other experimental group: thus, it was of definite assistance to have the possible function of the ceiling pointed out. If, then, subjects are furnished with an experience in order to ensure their having a given principle, it is necessary to ascertain whether or not this principle has, in fact, been grasped. The distinction between the particular and the general is one which Maier failed to take into account when discussing the sufficiency of pre-problem experiences.

Another instance of inadequate handling of pre-problem experiences occurs in a later experiment by Maier (1945). One group of subjects were shown how to build a structure similar to one of the pendulums and then afterwards given the problem of constructing a hat rack. To solve this problem it was necessary to clamp two poles together and wedge them between the floor and the ceiling so that the clamp could function as a hook on which to hang a hat and a coat. Seventy-two per cent of the subjects solved this problem. A second group was present only during the construction of the hat rack, although no mention was made of its possible functions. When this group was asked to list all the possible functions which the construction might have, only 24% mentioned its use as a hat rack. If I interpret Maier aright, he argues that, since these two groups had both seen the same construction, they had been exposed to equivalent stimulus situations. However, this may be, it seems clear that the two groups received markedly different instructions and so faced the situation with different attitudes. The first group was confined by the instructions to the concept of a hat rack. The second group was instructed to produce as many different concepts as possible. In no investigation of problem-solving can we afford to ignore the role played by the instructions given to the subjects (Wertheimer, 1945, seems to be guilty in this respect).

Limiting the present discussion to practical construction problems of the Maier type, it may be said that the subject has a goal which he must attain by using a definite number of objects in definite ways and that these objects serve definite 'functions' in the attainment of the goal. If problem-solving is thought of in terms of these 'functions', there arises at least one important question. When a problem is not solved, is it simply the result of one or more of the necessary 'functions' being unavailable to the organism, or may it also be due to some difficulty in combining these 'functions'? Or, otherwise expressed, if a subject has available the necessary 'functions' for the solution of the problem, will he actually solve it?

An answer to the above question demands that the availability of 'functions' be determined independently of the subject's behaviour in the problem-solving situation. This

determination may be made by presenting the subject with the objects which are later to be used in solving the problem and asking him to list all the 'functions' which he thinks each object could have. In this way, there may be obtained in advance subjects who have available the necessary 'functions' for the problem's solution. Which particular 'functions' are or are not necessary depends, of course, on the exact nature of the problem later to be set. Furthermore, there is no guarantee that a subject who fails to list a particular 'function' beforehand may not get the 'function' when presented with the problem situation. The main interest of the following experiment will therefore centre around the group of subjects who give, in advance, all the necessary 'functions'.

Before going on to describe the actual experiment, it should perhaps be indicated what the writer regards as the three main requirements of any problem situation in which the role of the availability of 'functions' is to be examined. First, there should be several objects and thus several 'functions' to combine. With too few 'functions' it might be argued that the situation did not constitute a fair test of the role played by the combination aspect of the solution. It is noteworthy that the problems usually investigated in this field rarely involve more than just a few 'functions'. Secondly, the problem must not be presented in such a way that the subject doubts, for some trivial reason, the appropriateness of an object to any necessary 'function'. If a rod, for example, has to be used for horizontal support, the objects to be supported must not be so heavy that the subject comes to doubt whether the rod is strong enough to carry their weight. Thirdly, if a time limit is imposed, it is important that the subject be able to discover fairly rapidly what erroneous modes of attack are, in fact, wrong. A wrong, but plausible seeming, approach may so engage the subject that he never attempts the correct solution of which he may well be capable. More than one correct solution might also complicate matters, for, having found one, especially if it has been found with difficulty, the subject will tend to rest content with it and not attempt the other.

II. EXPERIMENTAL

(i) *Subjects*

The subjects were 205 male college-student volunteers, aged 17–20 years. Of these, 120 took part in a pilot investigation designed to standardize the problem and the instructions. The experiment proper involved 85 subjects of whom 57 were allocated randomly to an Experimental Group and 28 to a Control group.

(ii) *The problem*

The problem was a modification of Maier's (1933) candle problem. A candle, about 5 in. high, had to be blown out from a distance of 70 in., the candle being placed on one table and the subject standing behind another. Available to the subject were: six hollow glass tubes, 9.5 in. long and $\frac{3}{4}$ in. in outer diameter; a solid duralumin rod, 50 in. long and $\frac{5}{16}$ in. in diameter; a 1 oz. lump of putty. The particulars of the materials used are of importance since the pilot study revealed that they influenced both the difficulty of the problem and the frequency of listing the various 'functions'. Each subject was given the following instructions. 'Here is a problem for you. I want to see if you can put out this candle from behind this table. I want a safe good method. If you do not feel like doing it, you may explain to me how you would do it. You may use this material [the experimenter

pointing to the material on the table] or not'. If the subject suggested an incorrect approach to the problem he was not informed that it was wrong but asked to perform what he had in mind. If he tried to lengthen the rod either by stretching across the table or by fixing a glass tube on to the end nearest him, he was told that one end of the rod had to be in line with the edge of the table nearest him. It might have been possible for the subject to solve the problem by lengthening the rod, through putting glass tubes on to its end, and then tapping out the flame. To prevent this, the cohesive power of the putty was lowered by moistening it slightly and the glass tubes were brought in wet after newly having been washed. The latter device also served to reassure the subject that the glass tubes were clean enough to be touched by his lips.

Under the above conditions, the only possible solution was to put the glass tubes end to end round the rod and, by means of the putty, make the joints between them air-tight. By kneeling down the subject could then lift the blow-pipe structure (preferably while leaning his elbows against the table), direct it at the flame and, by blowing through it, extinguish the candle.

The pilot study showed that not all the subjects who saw the solution were able to manipulate the practical details of it. The following criterion for the correct solution was therefore, adopted: the subject must place at least five tubes on the rod, seal up at least one joint completely, and either blow or clearly express an intention of so doing. Merely placing a lump of putty across the joint of two tubes was not considered sufficient since subjects often did this preparatory to sticking the tubes together in line rather than to sealing them into one long airtight tube. A subject might attempt to bring the end of the construction that was nearest to the candle to the same level as the flame by pressing down on the table with the end nearest him. In this case, he was told that the tube would break.

(iii) *Procedure*

Each of the 85 subjects was presented with the candle problem as described above. The time limit was 30 min. and if a subject wanted to give up before then he was encouraged to continue until the time was up. The Experimental Group differed from the Control in that it was given two previous sessions in which the availability of the necessary 'functions' was tested for.

The first session given a subject in the Experimental Group lasted for 15 to 30 min. and was designed to make clear to him what was required in the availability test proper. The necessity of including this session was made clear by the pilot study. The subject was told that he would be shown a number of objects and that he was to list all the 'functions' which he could think of in connexion with each. The objects were not to be thought of as changeable, i.e. as being alterable by, say, breaking or melting. When he had a 'function' in mind, he was to state, first the physical characteristics of the object which allowed it to serve this 'function' and, secondly, the 'function' in a general form giving three concrete examples. He was told to keep the listing of each 'function' separate from the listing of every other function. The practice object was an iron bolt. During this session, the subject was free to question the experimenter on any points of difficulty.

The availability test proper lasted for 13 to 30 min. The objects necessary for the solution of the candle problem were presented in the following order: the glass tube, the lump of putty, the rod. The three objects were shown one at a time and, for each, the subject

listed its 'functions' until he could think of no more. The 'functions' were stated verbally and noted down by the experimenter. Those 'functions' which would later be necessary for the solution of the candle problem were three in number: (1) the glass tube had to conduct an air stream when blown through, (2) the putty had to make crevices, apertures, joints, etc., either air- or water-tight, and (3) the rod must support objects hung from it when held in a horizontal position. A 'function' was judged as given when it was stated in a general form accompanied by one appropriate concrete example. The subject's statements relevant to the three necessary 'functions' were given to two psychologists who scored them independently of the experimenter and of each other. Subjects in the Experimental Group were given the candle problem immediately after completing the availability test.

(iv) *Results*

As stated in the Introduction, the experiment's chief interest was the performance of these subjects who, in the availability test, gave the three 'functions' of 'blow', 'make tight', and 'hang'. However, after testing 35 subjects in the Experimental Group, it was found that only one gave all three 'functions'. Yet every one of seven subjects who gave only the second and third 'functions' solved the problem. It was decided accordingly to continue the experiment by regarding 'make tight' and 'hang' as the necessary 'functions' for otherwise a very large number of subjects would have had to be run.

Of the 56 subjects in the Experimental Group¹, 13 (group 1) gave the functions of 'make tight' and 'hang'. All 13 solved the problem. There were 43 (group 2) subjects who failed to give both 'functions', and only 25 of these solved the problem. If the difference between those two groups, with regard to the number of solutions, is tested by Fisher's exact method, a *P* value of 0.003 is obtained on a one-tailed test of significance. Of the 18 subjects in group 2 who did not solve the problem, 11 did not use the rod for hanging and 7 placed the glass tubes on the rod and attempted to blow out the candle without sealing the joints.

The two psychologists who scored the statements disagreed on one statement relevant to the function 'hang' and on four relevant to 'make tight'. Only subjects on whose statements there was agreement were included in group 1.

The performances of the same 56 subjects were also assessed on the basis of a slightly less rigorous statement of the two necessary 'functions'. They were credited with the availability of both 'functions' if their comments merely implied that: (1) the putty might make tight in some way (e.g. tight against water, air, light, or insects; covering holes to obtain a smooth surface was not acceptable), and (2) the rod might act as a horizontal support. In terms of those criteria, 27 subjects (group 1) had both 'functions' available and 23 (group 2) had not. In group 1, the problem was solved by 23 subjects and in group 2 by only 11. If the difference between those two groups is tested by a χ^2 corrected for

¹ One subject was left out of the material as, after 10 min. he managed to blow out the candle without making the joints tight with putty, and was unwilling to continue, arguing that he had solved the problem. This subject's solution is a little doubtful. He placed five tubes on the rod and applied putty almost all round one joint, but not on the lower part of the joint. When asked after the session why he had not made the joints tight, he contended that that was what he had been trying to do, but there was no reason to seal on the lower part of the joint as no air would escape there. In contrast to two other subjects who came so far towards the solution that they placed a lump of putty over a joint, this subject maintained that he had intended to make the joints tight and not just to try to stick two tubes together by the putty. Putty was never used for glueing purposes, he said. Both his overt and verbal behaviour seem to be different from that of the other two subjects.

continuity, a P value of 0.006 is obtained on a one-tailed test of significance. (It will be noted that six subjects have been excluded from these results. This is because the two scorers could reach no agreement as to whether or not those subjects had given both necessary 'functions' in the availability test.)

Of the 28 subjects in the Control Group, 16 solved the candle problem. When this number of solutions is compared with that of the Experimental Group by means of χ^2 corrected for continuity on a two-tailed test of significance, the P value is 0.47. Thus there was no difference in the ability of the two groups to solve the problem. Nor was there any difference between the groups with regard to the time required for the solution: the successful members of the Experimental Group required an average of 10 min. 30 sec. for solution, and those of the Control Group an average of 10 min. 33 sec. Observation suggested, however, that there may have been qualitative differences in the problem-solving behaviour of the two groups. If the correct solution is regarded as a definite number of 'functions' performed in a definite sequence (i.e. blowing with the mouth, blowing through one tube, blowing through two tubes, putting the tubes on the rod and blowing, and then sealing and blowing), the Experimental Group seemed to perform in a less smooth sequence than the Control and with more trial-and-error. But a detailed discussion of those qualitative differences cannot be given here.

III. DISCUSSION

The above results seem to indicate (1) that subjects able to give the necessary 'functions' in the availability test are later able to solve the set problem, and (2) that the availability test, as such, may have little or no effect on the subsequent problem-solving behaviour. They suggest that the difficulty of any problem rests in the unavailability of one or more of its necessary 'functions' rather than in any difficulty of combining these 'functions' together. Further, it would not seem necessary that all the necessary 'functions' be combined at the one time; it may be that most problems can be broken down into sub-problems involving a smaller number of 'functions' than does the original.

The critical theoretical issue of the present paper concerns the essential difference between a subject who solves a problem and one who does not. Are they to be thought of as (1) differing in the availability of necessary 'functions', or (2) differing in the possession of an 'ability to combine functions'? In so far as the above experiment succeeded in establishing objective criteria for the necessary 'functions', for the availability of these 'functions', and for the solution or non-solution of the problem, the first alternative seems the more plausible. After all, there is, as far as the writer knows, no clear evidence of the existence of an 'ability to combine functions'. But if there were such an ability, it might be possible to explain the above results in terms of it. Thus it could have happened that the availability test selected, for group 1, subjects of superior 'verbal ability', and this ability might correlate highly with that required in 'combining functions'. In that case group 1 would comprise a selected group of subjects of superior 'ability to combine functions'. On the other hand, it would be equally admissible to argue instead that the 'verbal ability' correlates highly with ability to make 'functions' available. Another way in which an 'ability to combine functions' might vitiate the present interpretation is as follows: If the students participating in the experiment all had a high 'combining ability', then the only feature which would differentiate among them would be the

availability of the necessary 'functions'. This argument, however, appears to return us to the original interpretation. It might, of course, be that 'verbal ability' does not correlate too highly with either 'ability to produce functions' or with 'ability to solve problems'. This is perhaps suggested by the fact that when the criterion of availability was made less rigorous (i.e. when the subject need only imply a reference to the necessary 'functions') the results seemed little altered.

It should be made clear that the writer does not deny an 'ability to combine functions'. He merely asserts that such an ability has not yet been established, as some psychologists appear to imagine, and wishes to draw attention to it as a problem for future research.

So far we have avoided any discussion of what may be involved in the availability test. Did it make the combination of the 'functions' easier, or more difficult? As long as we know so little about what processes take place when someone produces a 'function' in the availability test, it is difficult to answer this question or even to state what is meant by 'combination of functions'. Comparison of the Experimental and Control Groups does not reveal that the availability test gave the Experimental Group any advantage over the Control with regard to solving the candle problem. It is, in fact, difficult to see how the procedure in the availability test could make any one combination of 'functions' easier than any of the other many combinations possible with regard to the numerous 'functions' of the three objects. Indeed, it is not impossible that, by making available a larger number of 'functions', the availability test might actually make the solution of the problem more difficult. There is, too, the possibility that a 'function' might be available during the availability test and yet, for some reason, be unavailable in the actual problem situation.

In concluding we may comment on the advantages to be gained if our interpretation in terms of availability of 'functions' is valid. In the first place, it would seem to imply that the number of 'functions' that have to be combined does not necessarily determine the difficulty of the problem. As long as the necessary 'functions' are available, it probably does not matter greatly whether there are few or many. This implication might help to increase our understanding of what is meant by the term 'problem' and also help to integrate psychological material which, at present, tends to be dealt with under such separate headings as 'learning', 'problem solving', 'memory', 'perception', 'motivation', and 'intelligence'.

Whether or not the present approach is fruitful is a matter of the usefulness of the questions which it raises. For example, are 'functions' combined by being added together in successive fashion? Does the solution develop as a systematic trying out of all available 'functions' until the correct pattern is obtained? Would it be correct to predict that, in any problem situation, the ability to produce 'functions' is highly correlated with the ability to solve problems?

IV. SUMMARY AND CONCLUSION

When a problem involves a goal to be reached by the manipulation of concrete objects, these objects may be said to serve certain 'functions' in the attainment of the solution. Eighty-five college students were set a modified version of Maier's candle problem. For 17 of these subjects the problem was preceded by an availability test. In this test nothing was mentioned of the problem itself, but the objects later to be used in it were presented one by one and the subject instructed to list all the possible 'functions' which the objects

might serve. By this technique, a group of 13 subjects was obtained who, in advance of the problem situation, had given evidence that the necessary 'functions' were available. All 13 subjects later solved the problem. Of the subjects who did not indicate that the necessary 'functions' were available, only 58% solved the problem. Stated in the most general terms, these results are interpreted to mean that a subject will solve any problem if he has available the necessary 'functions' and that it is not the difficulty of combining these 'functions' which renders a problem insoluble for a particular subject. The implications of this interpretation are briefly discussed.

The writer wants to express his indebtedness to his sponsor L. L. Thurstone, to E. H. Porter Jr., Austin H. Riesen, Garth J. Thomas, Donald W. Fiske and Lyle V. Jones, and to Dr I. M. L. Hunter (University of Edinburgh) for invaluable help with the manuscript.

REFERENCES

- HARLOW, H. F. (1951). Thinking. Pp. 452-505 in *Theoretical Foundations of Psychology*. (Editor, Harry Helson.) New York: Nostrand.
- HUMPHREY, G. (1951). Thinking. London: Methuen.
- MAIER, N. R. F. (1930). Reasoning in humans: I. On direction. *J. Comp. Psychol.* x, 115-43.
- MAIER, N. R. F. (1933). An aspect of human reasoning. *Brit. J. Psychol.* xxiv, 144-55.
- MAIER, N. R. F. (1937). Reasoning in rats and human beings. *Psychol. Rev.* XLIV, 365-78.
- MAIER, N. R. F. (1945). Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. Exp. Psychol.* xxxv, 349-60.
- WERTHEIMER, M. (1945). *Productive Thinking*. New York: Harper.
- WEAVER, H. E. & MADDEN, E. H. (1949). Direction in problem solving. *J. Psychol.* xxvii, 331-45.

(Manuscript received 28 May 1954)

A STUDY OF THE PROBLEM OF INTELLECTUAL DETERIORATION IN ADOLESCENTS AND YOUNG ADULTS

By K. LOVELL

Borough Road College, University of London Institute of Education

I. Introduction (pp. 199–200). *II. Outline of previous research* (pp. 200–202). *III. Plan of the present investigation* (pp. 202–205). *IV. Discussion and results* (pp. 206–208). *V. Summary* (p. 209). *Appendix* (p. 209). *References* (p. 210).

I. INTRODUCTION

Some decrease in intellectual efficiency is one of the commonest manifestations of mental disorder, cerebral lesion and age, while the term deterioration is generally taken to mean a progressive change in the direction of less and less efficiency. Moreover, it is usually implied in the literature that deterioration involves an irreversible change, although this is seldom stated explicitly. In this study it must be understood that the term is used in a more neutral sense, since we cannot be sure if the levels of performance found in the different groups taking part in this investigation could be improved upon given an environment more intellectually stimulating. Furthermore, it is impossible to say if the levels of performance found indicate a decline compared with those that would have been found earlier, or whether some abilities in certain of our groups never came to fruition.

The problem of intellectual deterioration among normal adolescents and young adults is one which has been studied but little. Psychologists and psychiatrists have certainly made great efforts to measure intellectual deterioration in populations suffering from certain organic and psychopathological conditions, but among younger, non-clinical cases, the problem has been largely ignored, while the findings among clinical cases to date have been somewhat confused and contradictory. This lack of clear findings is thought to be due in part to the failure of research workers to pay proper attention to the design of their experiments and to the basic principles of mental testing. Again, the dearth of experimentation among normal young people conducted along lines similar to those used in the case of clinical subjects is, presumably, due to the fact that it was considered that intellectual deterioration did not occur among the former. Now, however, it is clear that general intelligence begins to decline in a normal (i.e. non-clinical) population at a comparatively early age, and the rate of decline seems dependent both upon the original level of intelligence and upon occupation, while verbal educational abilities are certainly affected by teaching and environmental opportunities generally. Differences in level of performance in respect of some other abilities might therefore conceivably be found among groups with varying backgrounds. Furthermore, it is possible that such differences are the same as some of those said to be indicative of changes in efficiency associated with the onset of certain clinical conditions.

Vernon & Parry (1949) summarized the position as it was up to about 1947 in relation to the effects of age on intelligence test scores. Norris (1940), in an extensive investigation, found that educational attainments tend to lapse rapidly after leaving school except in so far as they are used in daily life. Again, Vernon & Parry (1949) suggest that verbal educational abilities (other than reading ability which shows an improvement) seem either to

remain stationary or to decline over the period 14 to 18 years of age unless they are stimulated by secondary or further schooling. As a result of comparing scores obtained by Glasgow school leavers and Glasgow naval recruits they submit that spatial mechanical abilities may rise over this period whether or not further education has been received.

Over and above these well-established g , v , k abilities, which are of such proved worth in education and in vocational guidance, there are many notions which have, in recent years, been discussed and developed in the literature concerning 'concept formation', 'rigidity' and the like. It is true that these terms have mostly arisen out of research with clinical cases of one type or another and the results are, to say the least, somewhat confused. Nevertheless, if factors of 'concept formation' or 'rigidity of thinking' exist, over and above g , v , and k , these should be investigated, since it is possible that deficits found among mental patients may also occur in young, non-clinical cases. Typical of the viewpoint that concept formation is inadequately measured by current tests of intelligence is that expressed by Hearnshaw (1951). After condemning the view of Burt and Vernon that the map of the intellect is more or less complete, he argues that he suspects that into the process of concept formation there may enter a type of abstraction that enters only incidentally into conventional tests of intelligence. 'These tests', writes Hearnshaw, 'involve for the most part simple relation and correlate education, not the ordering of manifold and heterogeneous experience through the grasp of abstract principles of order. Thus Matrices presents a determinate system of relations one term of which is missing. This is quite different from a determinable principle of order which can be variously exemplified. The Matrix is relationally concrete in comparison with the concept

II. OUTLINE OF PREVIOUS RESEARCH

The literature surrounding this study is extensive; mention can be made of only a few of the more important references. Goldstein & Scheerer (1945), Hanfmann & Kasanin (1937), Berg (1948), and others, have devised a number of concept formation or sorting tests. In a typical test the subject is presented with a series of objects which can be classified in various ways and he is told to sort them according to a given principle or pick out others similar to a given object. When the subject has mastered one principle of classification he is given a new and more complex problem. It is claimed that brain-injured patients and some schizophrenics can cope with these at the concrete level, but cannot realize and formulate abstract principles of classification, nor can they switch from one principle of classification to another.

A further concept to be considered is that of perseveration, the tendency to inertia in mental processes, as contrasted with the ability to switch quickly from one process to another. Walker, Staines & Kenna (1943), and Cattell (1946) now state that the earlier tests so often used to measure perseveration (e.g. the alternation type of test and miscellaneous sensory tests) are useless; but that 'creative effort' tests, where the subject has to break down some well-established habit, are more consistent and measure a quality which has been termed 'disposition rigidity'. Cattell claims that a person with low disposition rigidity is a person of strong ego development and capable of modifying his habits. Little, however, seems to be known of the validity of tests of disposition rigidity.

Luchins (1942, 1947) maintains that inefficient methods of education may produce rigid rather than flexible minds and has devised a number of paper-and-pencil tests

Subjects are given a series of problems which can be solved by the application of one principle. Half way through the principle alters, but the original 'set' may delay or prevent them entirely from realizing this. The interference in such tests is established temporarily during the course of the test, whereas in disposition rigidity tests the interference is with a fully automatized habit. Yet another concept of rigidity has been put forward by Lewin (1935) and Kounin (1941). These and other Gestalt-minded psychologists regard rigidity as one of the main characteristics differentiating the personalities of the feeble minded from those of normal children. When given a choice of drawing tasks, feeble-minded children tend to persist longer at a single task, whereas normal children prefer a change.

Again, there is a mass of literature dealing with tests of the type included in the Wechsler-Bellevue Battery (1944), the Babcock-Levy Battery (1940) and the Shipley-Hartford Test (1940). The performance of psychotics, especially those of the organic type, is poorer than expected on some of the tests; performance also declines in aphasic conditions, in cases of brain lesion and to some extent with age. The decline is least marked in tests of what Cattell calls 'crystallized' intelligence, such as vocabulary. Tests of 'fluid' intelligence, which are most affected, include speed tests and tests of abstraction or conceptualization.

Few attempts have been made, until the present inquiry, to determine whether tests such as those described by Goldstein & Scheerer, tests of disposition rigidity, and tests of the Wechsler-Bellevue Battery, in fact measure anything over and above the well-established g , v , k abilities. Absence of such work may be due to the following considerations:

(a) Most of the relevant work in this field has been done in America where there is less agreement on a small number of ability factors.

(b) Clinical workers, and those generally who have done most of the work in this field, seem often to be hostile to a factorial approach, or else ignorant of or indifferent to the value of such an approach.

(c) Since the available tests have been individual ones there would have been an enormous amount of labour expended in the giving of a battery of say concept formation tests to a group of a size adequate for a factorial approach. The one serious attempt made to use factor analysis was that of Halstead (1947) who applied nineteen tests to fifty patients. Neglecting the analysis which led to an oblique factor structure and concentrating on the analysis performed by Holzinger which gave a general and orthogonal group factor solution, it seems clear that the general factor is general intelligence or g , while Halstead's A factor is possibly the ability to group to a criterion for it has its highest loading 0.6 on a category test. This work has been criticized by Vernon (1950).

Crown (1951), in a review of the literature dealing with psychological changes following prefrontal leucotomy, comments on the lack of consistency between investigations on the issue of whether or not abstraction is affected following the operation. This lack of agreement is not surprising, he maintains, considering the variety of tests which have been used to measure abstraction. He divides these tests into two main types:

(1) Those which are primarily intelligence tests—Kohs Blocks, vocabulary tests, Shipley, proverbs, fables and abstract words.

(2) Tests which appear to conform to Halstead's definition of abstraction as the basic capacity to group to a criterion. Tests of the latter type involve the comprehension of essential similarities and differences—as in, say, the Goldstein-Scheerer Colour Form

Sorting Test. 'It is suggested', writes Crown, 'that, in factorial terms, sorting tests, if included in a test battery with other cognitive tests, would cluster together, these tests having, in addition, moderately high correlations with intelligence. Whether the abstraction group factor would be different from group factors already widely recognized remains a problem for further research.' In the opinion of the writer Crown's suggestion comes remarkably near the truth.

III. PLAN OF THE PRESENT INVESTIGATION

The general plan of attack on the problem of intellectual deterioration in adolescents and young adults adopted by the writer was as follows. A review was made of a representative selection of the literature dealing with the ability to think abstractly, to conceptualize and to switch from one conceptual set to another in psychotics, organics, neurosurgical cases and normals. Linked with this was a study of the literature dealing with motor and cognitive rigidity. From the overall field a selection of tests was made which appeared to be best substantiated by past research. To these tests was added a test of non-verbal intelligence and one or two of the more usual tests used in educational research, i.e. reading and arithmetic. The test battery was applied to various groups of young people, all non-clinical subjects, between the ages of 14 and about 20 years, and the results submitted to factor analysis to discover whether in addition to *g*, *v:ed* and *k:m* abilities any further group factors could be found such as:

- (a) ability to reason at the abstract or concrete level.
- (b) ability to switch from one conceptual set to another.
- (c) ability to switch from one set of habits to another in the motor or cognitive fields

Using point biserial correlations in connexion with group differences it might be possible to validate any additional factors found, against an external criterion.

The experimental population consisted of an older and a younger group. The testees in the latter category were still at school and were in turn divided into two groups, one of which came from a co-educational Secondary Modern School where the atmosphere was considered to be stimulating. This group consisted of boys and girls in their last year at school. The other group, which consisted of boys only, also in their final year, came from a Secondary Modern School where the atmosphere was reckoned to be non-stimulating. It is admitted at once that the criterion of stimulation was based on subjective judgements but the judgements were made without dissent, by Training College tutors who had had good experience in visiting different types of school. To be specific we may say the school which provided a stimulating atmosphere was characterized by:

- (a) a good standard of teaching maintained by the staff employing in the main a formal approach to teaching;
- (b) A good attitude to work on the part of the pupils whether one considered the more academic subjects or, say, music or craft;
- (c) An excellent spirit existing between the children and the staff.

It is a truism to say that the school often reflects the neighbourhood, but this point must be stressed. In areas where the children come from good stock and helpful homes there also the more able teachers tend to congregate and stay. As a result, the life and work of the school, already being built on potentially sound material, are further enriched. Thus it is quite impossible to disentangle the school and environmental influences, and we must

leave our discussion of the degree of stimulation afforded by the background of our school groups by saying that the stimulated group not only came from a better school, but most probably from an intellectually healthier environment and more helpful homes, and that the children within the group were rather more stable in their behaviour. The average age of both of the school groups was close to $14\frac{1}{2}$ years, and it must be emphasized that the economic conditions within the groups was about the same as far as could be judged.

In the other half of our experimental population there was, once again, a stimulated and non-stimulated group. The former group was made up of young men and women who were working in factories, offices or at trades, and who came to a Day Release Centre for one day per week. The members of this group had the following characteristics:

(a) They were all employed by progressive firms or organizations that made attendance at a Day Release Centre a condition of employment;

(b) They attended a Day Release Centre where the standard of teaching was assessed as good, where the attitude of the students to work was sound and where staff-student relationships were excellent.

(c) They had been exposed to the influence of work and of the Day Release Centre for as long as possible before testing.

(d) They were not the most able people attending the Centre (our purpose was to get a group comparable in ability to our school groups).

The very fact that this section of our population was holding down posts with progressive firms or organizations would indicate that it was composed of people who had some liveliness of mind and some stability of behaviour, for the bad work record of the delinquent and maladjusted is well known. The average age of this group was $17\frac{1}{2}$ years.

The problem of getting an older non-stimulated group is a formidable one, principally because normally one has no hold over such a group which would make it submit to a testing programme that lasts several hours. Several possibilities were explored. Eventually through the kindness of the Prison Commissioners¹ the difficulty was overcome, permission being given to test a group of young prisoners at a rather special type of prison in the south of England that has been set aside for such offenders. To be eligible to serve a sentence there the prisoner must be under 21 years of age when sentenced, and the sentence must be for a period not greater than four years. Our request was for a batch of prisoners who had been admitted to the prison within a few weeks of the opening of the testing programme. Thereby it was hoped that it would be possible to examine these young men literally as they 'came off the streets', and before prison work and routine had had any appreciable effect on them (some, of course, had been to an Approved School or Borstal Institution or both). A further request was made that the outstandingly able, the very dull and all illiterate prisoners be excluded from the test group, as it was hoped that it would be possible to make the group comparable with the other groups as far as ability was concerned. Our total population thus consisted of four groups of approximately fifty subjects in each group, making some 200 cases for examination in all.

The conditions governing the choice of tests were as follows:

(a) The tests had to be representative of the many different types that have been used in the field of mental deterioration, and at the same time they had to be among those that appeared to be most promising.

¹ Any opinions expressed do not necessarily represent official views.

(b) It had to be possible to put all the tests into group form. The Vigotsky Blocks Test was ruled out on this account.

(c) The overall testing time, including time devoted to reading out the instructions and ensuring that every one of the testees understood what had to be done, could not be more than about eight hours.

(d) Bearing this maximum in mind it was essential to devise group versions of the chosen tests which would yield reasonable reliabilities, and which might therefore obtain reasonable communalities in the factor analysis.

When appropriate tests had been selected or prepared, all, apart from well-known tests, were tried out on the older age groups of a Secondary Modern School (not one of the two schools from which the experimental groups came). In some of the cases the time allowances, level of difficulty of the items, etc., seemed to be about right on the first trial, but in other cases considerable alterations had to be made and the revisions tried out again on different but comparable children. Although it had been hoped that the trial groups of children were intellectually comparable with the experimental groups, it is the writer's opinion that the latter were more homogeneous with respect to ability than the former. This means that our correlations are somewhat lower than those to be expected from a more representative sample. The tests finally included are now listed with a few relevant comments:

Test 1. N.I.I.P. Group Tests 70/2 and 70/3.

Test 2. Watts-Vernon Reading Test.

Test 3. Arithmetic Test. This consisted of fifty problems similar to the problem section of the Moray House type of Arithmetic Test and thus similar to the Wechsler-Bellevue Arithmetic Test, but much longer.

Test 4. A shortened version of the Wechsler Vocabulary Test.

Test 5. A group version devised by the writer of the Wisconsin Sorting Test. This test was scored in two ways—for the principle of classification ('labels'), and for sorting the pieces correctly ('pieces').

Test 6. A group version devised by the writer of a sorting test suggested by Vinacke (1951). This test was scored in two ways as in the case of Test No. 5.

Test 7. A group test prepared by the writer based on the Trist-Hargreaves test (Semeonoff, 1955). This also was scored for the principle of classification as stated, and for the pieces correctly classified.

Test 8. A test of Concept Formation of the Heidbreder (1949) type.

Test 9. A non-verbal Classification Test. The test was scored for the underlining of the correct figures and for stating the principle of classification.

Test 10. Proverbs.

Test 11. Test of Memory for Stories.

Test 12. Babcock Test of Memory for Sentences.

Test 13. Test of Memory for Designs including fourteen of the designs used in the Bender Visual Gestalt and Ellis Tests.

Test 14. Creative Effort Test. The writing of ZEK forward and then in reverse as ZEK.

Test 15. Creative Effort Test. The writing of 239 forward and then in reverse.

Test 16. Creative Effort Test VAT. The letters were written first in the normal way and then upside down.

Test 17. A Creative Effort Test in which there were 46 very simple addition and subtraction exercises but whenever there was a plus sign the testee had to subtract and vice versa.

Test 18. Similar to Test 17 but consisting of interchanged multiplication and division exercises.

Test 19. Creative Effort Test—Seasons. Here the subject had to associate a month of the year with a season, e.g. August with summer. But sometimes the process had to be reversed, e.g. April had to be associated with autumn.

Test 20. This was the first of three tests which in themselves are entirely original although of the type used by Luchins. It involved the simplification of very easy fractions and was in two parts the first of which measured basic speed of working.

Test 21. Hidden Words. The test was in two parts, the first of which measured the basic speed of working.

Test 22. Beginnings and Endings of words. Again the test was in two parts, the first measuring the basic speed of working. In each of tests 20, 21 and 22 there was an attempt to develop a mental set in Part I of the test and to measure the extent to which this set interfered with the working of Part 2.

Test 23. Three different operations were first performed separately and afterwards similar operations had to be performed in quick rotation. It was suggested by the Army Clerical Test (Vernon & Parry, 1949).

Test 24. Gottschaldt Figures: 42 items from Thurstone's revision.

Test 25. An extension of the Shipley Test made by Vernon.

Test 26. A group version of the Kohs Blocks Test devised by Hey (1951).

In the case of Tests 14–16 and 20–23, basic speed scores were graphed against the rigidity scores (i.e. the scores obtained in the second part of the first six tests listed above, and in the fourth part of Test 23) and the approximate regression line drawn in. The rigidity of a particular case was then read off by subtracting from the obtained rigidity score, the rigidity score indicated by the regression line corresponding to the basic speed score. Rigidity was thus expressed in terms of a positive, zero or negative quantity and was made independent of the basic speed score. This appears to obviate criticism often levelled against earlier workers that the estimate of rigidity was dependent on basic speed; it also appears to solve a difficulty which Luchins and his followers seem to ignore.

All raw scores were first recast so as to make the distributions normal and product-moment correlations calculated after allowing for differences in subgroup means. Thurstone's Centroid Method of Factor Analysis, with subsequent rotation to a general and orthogonal group factors solution, was the method of factor analysis considered to be most appropriate to this experiment, while every effort was made to establish the well authenticated g , v and k abilities first. Finally, Point Biserial correlations were calculated in the case of each test for Stimulated versus Depressed or Non-Stimulated groups and for Youth versus Age groups. These correlations, however, were not included in the matrix; instead, correlations of 'Stimulation' and 'Youth' with the found factors were later calculated.

IV. RESULTS AND DISCUSSION

An examination of the means of the raw scores for the various subgroups revealed at once that the Stimulated groups did relatively much better on some tests than on others, and it was further established that this was not due to sex (it will be recalled that there were females in the Stimulated but not in the Depressed groups). For example, the Stimulated groups had relatively better scores on the Sorting Tests than on the Reading Comprehension Test. Accordingly, all group differences were reduced to standard scores. On adding together the mean standard scores of the two Stimulated groups and subtracting from this total the sum of the standard scores obtained by the Depressed groups (i.e. for each test in turn) it was seen at once that there was most deficit on the six concept formation variables and least on tests of reading, vocabulary and Group Test 70/2. A few results are given below; they strongly suggest at least that a form of deterioration may be present in the concept formation type of test not normally detectable in *g* and *v:ed* tests:

Test	Difference between stimulated and depressed groups	Test	Difference between stimulated and depressed groups
Three Activities	0.80	Kohs Blocks	-0.20
Wisconsin Pieces	0.69	Shipley Abstraction	-0.43
Vinacke Labels	0.62	Wechsler Vocabulary	-0.59
Wisconsin Labels	0.58	Reading Test	-1.01
Test of Fractions	0.62	N.I.I.P. 70/2	-1.06

The results of the factor analysis are set out in the Appendix with the factor loadings correct to two decimal places and with loadings less than 0.14 omitted. Six of the seven rotated factors obtained are, in the opinion of the writer, as clear-cut and unambiguous as can be expected from a factorial study involving tests in a new field. The first factor can be described as *g*, or general cognitive capacity, and without doubt indicates the presence of an ability which is responsible for most of the identifiable variance in non-verbal intelligence tests such as Matrices and Dominoes. The second factor is indicative of verbal ability; an ability of proven worth in educational and occupational prediction. The highest saturations of this factor occur on the Watts Vernon Reading Test, the Wechsler Vocabulary Test and on Proverbs. The third factor is clearly the spatial factor *k*; it correlates most highly with Memory for Designs, Gottschaldt Figures and the Kohs Blocks Tests. With the fourth factor this research appears to be breaking new ground. That is to say, factor four seems to be indicative of an ability long suspected by clinicians to be present in non-clinical cases but absent, in varying degrees, in certain clinical cases (i.e. brain lesion cases, organic psychotics and some functional psychotics). The factor has its highest loadings on the Sorting Tests, the test of non-verbal categorization or classification, and the three Luchins type tests. Note that the actual maximum loadings were obtained by tests 9B and 6L respectively, where in both instances, the subject has to say precisely how he has made the groupings. It is suggested that this factor should be labelled *c* for categorization (cf. Goldstein's notion of the categorical attitude) and that the core of the ability indicated by this factor seems to be the ability to group objects or ideas according to a criterion and to switch from one criterion to another. There is therefore a definite element of flexibility linked with this ability; indeed it is likely that the *c* factor is rather more complex than our analysis was able to bring out.

The fifth factor has been labelled non-perseveration (motor), since it has its highest

loadings on motor perseveration tests of the creative effort type. It is suggested that this factor may also be indicative of Cattell's disposition rigidity. Factor six is indicative of what Vernon (1950) has described as a numerical educational ability (*n:ed*). Summarizing, it may be said that the analysis has yielded the well-established *g*, *v*, *k* and *n:ed* abilities, a non-perseveration (motor) factor, and most important, it has yielded evidence of a factor indicative of the ability to categorize (independent of these other five abilities), long suspected but not previously demonstrated to the satisfaction of British psychologists.

A major criticism often levelled against the work of factorialists is that the found factors often seem to be little more than statistical artefacts devoid of any 'real' existence. In an attempt to overcome this criticism an effort was made to validate our analysis by correlating the variables of Stimulation and Youth with the found factors. The variable Youth is easy to interpret, it represents the difference in scores between the younger groups (i.e. the two school groups) and the older groups (Prisoners and Day Release students). The variable Stimulation, on the other hand, cannot be defined with such exactitude. There is no evidence that it is covered completely by the quality of the school teaching that the two groups (i.e. the stimulated and depressed groups) received. This may have contributed something but it is thought that the general background, attitude to life and work and the general adaptability of the two groups differed considerably. Thus the Stimulation variable will depend upon inherited traits, structure of the central nervous system (in the sense of being able to withstand a certain amount of stress and make adequate responses to real life situations), early upbringing and environment, cultural background, quality of teaching and school atmosphere generally.

The correlations between the two criteria and the six factors are now given with coefficients less than 0.14 again omitted:

	<i>g</i>	<i>v</i>	<i>k</i>	<i>c</i>	non- <i>p</i>	<i>n:ed</i>
Stimulation	—	-0.18	-0.24	0.50	—	—
Youth	0.17	-0.25	—	—	—	—

The following points emerge:

(a) There seems to be no relationship between *g* and the Stimulation variable. In other words, the stimulated and non-stimulated groups were matched for general intelligence.

(b) The non-stimulated group seems to have slightly higher verbal ability.

(c) The non-stimulated group did better on tests involving a large amount of *k* ability. This is precisely what one would expect for it has often been pointed out that delinquents do relatively better on *k* tests than do normals; also the stimulated group contained females, whereas the non-stimulated group did not.

(d) The substantial correlation between the Stimulation variable and the *c* factor provides strong evidence that this factor is not merely a statistical artefact. When it is remembered that our criterion cannot be a 'pure' one, the size of the correlation between the factor and the variable appears very encouraging. It suggests that our non-stimulated or depressed group is behaving very much like Goldstein's brain-injured patients.

(e) Verbal ability tends to rise between 14 and 20 years of age largely because vocabulary and reading ability increase. This confirms the data presented by the Ministry of Education (1950).

(f) There appears to be a positive correlation between *g* and Youth. But great care must be taken in suggesting that *g* deterioration has set in in our groups by about 15 years

of age or soon after. It is possible that our younger groups were really more intelligent than our older groups when the latter were $14\frac{1}{2}$ years of age. But, if the younger groups were considerably more able than the older groups had been at an earlier age then a smaller negative correlation between Youth and v would have been expected.

Another factor analysis was carried out by the writer in which the g , v and k factors were first removed from the complete correlation matrix, and the remaining intercorrelations factor analysed using Thurstone's Centroid Method with subsequent rotation of axes to yield orthogonal group factors. This method generally confirmed the above results; as did the findings of two other analyses made independently by Prof. Vernon using twenty-three of the variables.

It is submitted then that strong evidence has been presented indicating that the ability to categorize is not adequately measured by current g , v , k tests; also that the ability is present and absent to some extent in non-clinical as in certain types of clinical cases. But two further points must be made:

(1) The variance contributed by the c factor is small. In other words, even in tests which have the highest loadings on the c factor, g contributed in most cases, even more to the test variance. When Sorting Tests are given individually, as seems usual in clinical work, it is possible that the tests are easier than those used in this experiment and thus the tests are less dependent on g . In such a case the deficit on the c factor would show up more clearly since the factor would account for more of the variability in scores.

(2) It is impossible to know at the moment whether the relative inability of our depressed group to categorize is due to the fact that the ability never developed, or whether it was present in greater measure and has since declined. Again, as has already been said, we cannot be sure that the deficit in the ability to categorize is due merely to the quality of the school teaching and cultural background generally, or whether it is, as Goldstein would have us believe, symptomatic of a change in the response of an individual to the environment.

In an attempt to test the hypothesis that the ability to categorize can be improved by intellectual stimulation a further group of young prisoners was tested. In this case each of the testees had served at least nine months in the prison during which time each prisoner had had to attend at least one evening class per week lasting $1\frac{1}{2}$ hours and do one evening's Physical Training per week, quite apart, of course, from the normal work. There were also library facilities. Thirty such cases were matched against thirty prisoners in the first group for Matrices score, score on N.I.I.P. Group Test 70/23 and for Approved School and/or Borstal experience. Thus it was considered that we had thirty pairs matched for general intelligence. The scores obtained by the two groups on the Reading Test, Wechsler Vocabulary Test, Wisconsin Sorting Test, Vinacke Sorting Test, non-verbal Categorization Test, Proverbs Test and the Test of Three Activities were then studied, but the differences between the groups were not statistically significant except in the case of the non-verbal Categorization Test. On this test the second group of prisoners did rather worse than the first group. It is not possible to say at the moment, whether by intensive intellectual stimulation, the deficit could be made good.

Finally, it should be noted that our c factor contributes some of the variance of the Luchins type test. If this is confirmed by later research then it might be that the rigidity of Luchins is the same as the rigidity of Goldstein. As has already been said, however, the c factor is probably more complex than our analysis was able to show.

V. SUMMARY

1. There is an extensive literature surrounding the problem of intellectual deterioration in clinical cases. The results of past work are often difficult to interpret, due in the main to poorly designed experiments.

2. From an educational point of view we need to know if terms like 'conceptualization', 'rigidity of thinking' and so forth involve abilities other than g , v , k abilities.

3. Strong evidence has been presented indicating that there is an ability to group ideas or objects to a criterion and switch from one criterion to another not adequately measured by g , v , k tests.

4. Subjects in our non-stimulated group were unable to categorize as well as those of the stimulated group and in this sense behaved like Goldstein's brain lesion cases.

5. Whether a relative inability to categorize is due to poor-quality education and limited cultural background or whether it reflects a change in the response of an individual to the environment is a problem for further research.

6. The extent to which the discovery of our c factor will help in educational, occupational or clinical prediction is not yet known. The ability to categorize is probably more complex than our experiment was able to show.

The writer readily acknowledges the help that he received from Prof. P. E. Vernon in designing this experiment.

APPENDIX

Test	g	v	k	c	non- p	$n:ed$
1. N.I.I.P. 70/2	0.51	0.24	0.32	—	0.17	—
N.I.I.P. 70/3	0.54	0.20	0.25	—	—	—
2. Watts Vernon Reading	0.21	0.79	0.15	—	—	—
3. Arithmetic Problems	0.35	0.18	—	0.18	—	0.70
4. Wechsler Vocabulary	0.29	0.72	0.16	0.15	—	0.14
5L. Wisconsin Labels	0.58	0.31	—	0.26	—	—
5P. Wisconsin Pieces	0.56	0.30	—	0.28	—	—
6L. Vinacke Labels	0.51	0.22	—	0.43	0.15	—
6P. Vinacke Pieces	0.51	0.18	—	0.32	0.14	—
7A. Trist-Hargreaves Statements	0.66	0.26	—	0.18	—	0.22
7B. Trist-Hargreaves Pieces	0.58	0.17	—	0.34	—	0.17
8. Heibdreder Concept	0.37	0.46	—	0.23	—	0.25
9A. Non-verbal Categorization, Figures underlined	0.27	—	0.25	0.37	0.17	—
9B. Non-verbal Categorization, Reasons	0.44	—	0.24	0.44	—	—
10. Proverbs	0.39	0.61	—	—	0.26	—
11. Memory for Stories	0.30	0.56	0.26	—	0.19	—
12. Babcock Test of Memory for Sentences	0.21	0.44	—	0.16	—	—
13. Memory for Designs	0.29	—	0.62	—	-0.15	0.16
14. Creative Effort ZEK	0.15	—	0.20	0.14	0.53	—
15. Creative Effort 239	0.18	-0.20	—	—	0.57	—
16. Creative Effort VAT	—	-0.18	—	—	0.17	—
17. Plus and minus reversed	0.45	—	—	—	0.25	0.64
18. Multiplication and division reversed	0.37	—	—	—	0.14	0.75
19. Seasons reversed	0.38	0.15	—	0.16	—	0.40
20. Fractions	0.32	—	—	0.22	-0.21	0.18
21. Hidden Words	—	0.23	0.16	0.27	—	—
22. Beginnings and endings	-0.25	0.25	—	0.41	0.19	0.27
23. Three activities	—	0.16	—	0.29	0.34	0.16
24. Gottschaldt Figures	0.40	—	0.60	0.19	0.19	—
25. Shipley Abstraction	0.65	0.24	0.17	—	—	0.24
26. Kohs Blocks	0.45	—	0.62	0.19	—	—
Percentage variance	15.4	9.5	5.5	5.6	3.8	6.3

REFERENCES

- BABCOCK, H. & LEVY, L. (1940). *Revision of the Babcock Examination for Measuring Efficiency of Mental Functioning*. Chicago: Stoelting.
- BERG, E. A. (1948). A simple objective technique for measuring flexibility of thinking. *J. Genet. Psychol.* XXXIX, 15-22.
- CATTELL, R. B. (1946). The riddle of perseveration. 'Creative Effort' and disposition rigidity. *J. Personality*, XIV, 229-267.
- CROWN, S. (1951). Psychological changes following prefrontal leucotomy: a review. *J. ment. Sci.* XCVII, 49-83.
- GOLDSTEIN, K. & SCHEERER, M. (1945). *Tests of Abstract and Concrete Thinking*. New York: Psychol. Corporation.
- HALSTEAD, W. C. (1947). *Brain and Intelligence*. Chicago: University of Chicago Press.
- HANEMANN, E. & KASANIN, J. (1937). A method for the study of concept formation. *J. Psychol.* III, 521-40.
- HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.* XLII, 315-21.
- HEIDBREDER, E. (1949). The attainment of concepts. VIII. The conceptualization of verbally indicated instances. *J. Psychol.* XXVII, 263-309.
- HEY, C. G. (1951). M.A. Thesis, University of Birmingham Library.
- KOUNIN, J. S. (1941). Experimental studies in rigidity. I. *Character & Pers.* IX, 251-72.
- LEWIN, K. (1935). *A Dynamic Theory of Personality*. New York: McGraw Hill.
- LUCHINS, A. S. (1942). Mechanization in problem solving. The effect of Einstellung. *Psychol. Monogr.* LIV.
- LUCHINS, A. S. (1947). Proposed methods of studying rigidity of Behaviour. *J. Personality*, XV, 242-6.
- MINISTRY OF EDUCATION (1950). *Reading Ability*. London: H.M.S.O.
- NORRIS, K. E. (1940). *The Three R's and the Adult Worker*. McGill University.
- SEMEONOFF, B. (1955). *A Manual of Diagnostic Performance Tests*. In preparation.
- SHIPLEY, W. C. (1940). A self administering scale for measuring intellectual impairment and deterioration. *J. Psychol.* IX, 371-7.
- VERNON, P. E. & PARRY, J. B. (1949). *Personnel Selection in the British Forces*. London: University of London Press.
- VERNON, P. E. (1950). *The Structure of Human Abilities*. London: Methuen.
- VINACKE, W. E. (1951). The investigation of concept formation. *Psychol. Bull.* XLVIII, 1-31.
- WALKER, K. F., STAINES, R. G. & KENNA, J. C. (1943). P tests and the concept of mental inertia. *Character & Pers.* XII, 32-45.
- WECHSLER, D. (1944). *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins.

(Manuscript received 12 July 1954)

ADAPTATION TO LEVEL OF DIFFICULTY IN INTELLIGENCE TESTING

By A. W. HEIM

Psychological Laboratory, University of Cambridge

I. *Introduction* (pp. 211–212). II. *Test Material* (pp. 212–213). III. *Experimental design* (pp. 213–214). IV. *The subjects* (pp. 214–215). V. *Results* (pp. 215–220). VI. *Discussion* (pp. 220–221). VII. *Summary and conclusions* (pp. 221–222). VIII. *Acknowledgements* (p. 222). *Appendix* (p. 223). *References* (p. 224).

I. INTRODUCTION

There is some evidence that people unwittingly adapt their response to set tasks in accordance with the degree of difficulty presented. Naturally, individual variation obtains with respect both to this flexibility and to the type of response made. But in general people tend to 'rise to the occasion' in stimulatingly difficult circumstances and to 'sink to the occasion' in circumstances which present little or no challenge, giving then a mediocre performance on a relatively easy task. From this poor showing, the inference is sometimes drawn—perhaps invalidly—to a necessarily still poorer performance on a harder task.

Everyday life offers many illustrations of this type of adaptation. For example, players of team games are liable to excel themselves when confronted with formidable opponents and, conversely, to 'let the side down' when faced with a weak opposing team. The phenomenon has been displayed also in experimental situations. Mace (1931) demonstrated it in sensori-motor skills; Gibbs (1951) and others, in research on transfer of training; and Mowbray (1953) produced a very similar effect in his work on visual and auditory stimuli. Again, Heim's experiments (1949, 1950, 1954*b*) on repeatedly testing subjects with the same intelligence test suggested that the process has a part to play in mental testing. She found that a group of subjects of no more than average intelligence may produce a surprisingly high score if presented with a very high-grade intelligence test, believed on prior grounds to be 'too difficult' for them; whilst a group of the same intellectual level will fail to gain 100% on a much easier, and shorter, test of intelligence.

This suggests that the group tends to maintain a certain constancy of *relation* between performance and task, although the *absolute level* of performance may be less constant than is often supposed. It is clear that the phenomenon occurs only within certain limits: an impossible task, for instance, can scarcely evoke 'upward adaptation' from the group members who attempt it. On the other hand, the role of limits is not simple since, as is demonstrated in this paper and others, the limits themselves may shift as the subjects become upward- or downward-adapted (Heim, 1954*c*).

As already indicated, experiments on adaptation, in this sense, are numerous: though the terminology is far from uniform, studies on substantially the same problem are not lacking. But with the exception of the cited experiments on intelligence tests (and these were not primarily concerned with adaptation), the writer knows of no work in the field of mental testing which deals with adaptation to level of difficulty, either as the central

issue or as a factor to take into account when interpreting results. On these grounds, the present investigation was initiated.

Intelligence test problems of known difficulty were incorporated into contexts of varying levels of difficulty; and these mixed tests were administered to groups of subjects previously equated on intelligence test performance. By these means it was hoped to shed some light on the extent and the mechanism of adaptation to level of difficulty, in the field of intelligence testing.

II. TEST MATERIAL

The test items used in these experiments were drawn from the second parts of tests AH 4 and AH 5 (Heim, 1947). On the basis of item analyses of these tests (Wallace, 1952; Watts, 1953), data were available concerning the relative difficulty of their individual problems. The thirty-six diagrammatic questions comprising part II of AH 5 were ranked in order from 1, the hardest (that is, the one yielding the smallest percentage of correct answers) to 36, the easiest. Nos. 1-12 were, arbitrarily, categorized as 'difficult AH 5'; nos. 13-24, as 'medium AH 5', and nos. 23-36, as 'easy AH 5'. The threefold classification was made in order to facilitate the task of producing four tests, differing in the number of AH 5 questions included, but roughly similar in respect to the difficulty of these AH 5 questions.

Four mixed tests (A 1, A 2, B 1 and B 2), each comprising forty items, were then constructed, using the 36 classified AH 5 questions and the easiest 44 AH 4 diagrammatic questions (all of which were easier than any of the AH 5 questions). These mixed tests will hereafter be referred to as 'AH 45 tests'. Selected equated groups were to take the two A tests and other equated groups were to take the two B tests. All the items in the A series appeared also in one or other of the B tests; but no A 1 items were common to A 2 nor were any B 1 items common to B 2, since one or other pair of tests was to be taken by the same subjects.

Table 1 shows the distribution of items among the four tests.

Table 1. *Distribution of AH 4 and AH 5 questions among tests A 1, B 1, B 2 and A 2*

		A 1		A 2		Total
		AH 5	AH 4	AH 5	AH 4	
B 1	{AH 5	—	—	14	—	14
	{AH 4	—	16	—	10	26
						40
B 2	{AH 5	6	—	16	—	22
	{AH 4	—	18	—	—	18
						40
Total		6	34	30	10	
		40		40		

Thus for example A 1, the easiest of the four tests, consisted of 34 AH 4 items and 6 AH 5 items; these 6 AH 5 and 18 of the AH 4 items appeared also in B 2, the remaining 16 AH 4 items appearing in B 1. It may be seen that the ascending order of difficulty of the four tests is as follows: I, A 1 (6 AH 5 items); II, B 1 (14 AH 5 items); III, B 2 (22 AH 5 items); IV, A 2 (30 AH 5 items).

By means of the threefold classification of AH 5 problems, these were, in accordance with their difficulty, divided among the four 'mixed' tests as evenly as possible. For instance, of the six AH 5 problems in A 1, two were 'difficult', two were 'medium' and

two were 'easy'. When the four tests had had their items allocated, these were arranged cyclically so as to avoid consecutive series of AH 5 (or AH 4) questions, as far as possible. In test B 1, for instance, roughly every third item was an AH 5 question. The *type* of question too was distributed as evenly as was consistent with this plan, that is, a superimposition might be followed by an analogy and an analogy by a series—not another analogy. There was thus no gradient of difficulty in any of the four tests. The level of difficulty of adjacent problems varied markedly and, to all appearances, erratically. The only factor in common among all test questions was the form in which they were expressed: this was diagrammatic, the number of words used being at a minimum.

The decision was made to work with two pairs of tests rather than one pair, in order to ascertain whether the greater 'figure-ground' difference produced a greater effect, that is, whether adaptation to level of difficulty—in the event of its occurring at all—appeared more strikingly between tests A 1 and A 2 (the easiest and the hardest) than between (the intermediate) tests B 1 and B 2.

Time limits for the four tests were determined in accordance with the mean times which had been spent by university students on AH 4 and AH 5 questions, when taking the tests in unlimited time, as part of an earlier experiment (Heim, 1947). Eighty per cent of the mean time taken was allowed. When the time limits thus obtained were rounded off to the nearest minute, this method yielded the following set of time limits: A 1, 10 min.; B 1, 15 min.; B 2, 20 min.; and A 2 25 min.

III. EXPERIMENTAL DESIGN

Four groups of equated subjects were to take respectively A 1 followed by A 2, A 2 followed by A 1, B 1 followed by B 2 or B 2 followed by B 1. Thus the total testing time for all subjects on their pair of mixed tests was to be 35 min. Preliminary examples were, regretfully, dispensed with, as they might well have produced some unmeasurable adaptation to level of difficulty.

Some prior testing was necessary, however, since it was essential to equate the four groups on intelligence test performance. Thus all the subjects had in fact had previous experience on at least one other test, which was somewhat similar to the AH 45 series. The test selected for this purpose was Penrose's Pattern Perception (1944, 1946). This was chosen because it is a high-grade test intended for use with subjects of university status and it also has a diagrammatic bias. It comprises eight preliminary examples and sixty-four test items. These are classificatory in kind, the subject in every case being presented with five diagrams and required to select the 'odd man out'. A time limit of 30 min. (exclusive of the examples) is imposed.

On the basis of their Pattern Perception scores, the total group was to be divided into four subgroups with similar means and variances. The time interval between the Pattern Perception and the AH 45 testing sessions ranged from 4 to 7 days. Subjects were naturally not informed of the purpose of the inquiry until they had completed both testing sessions. They were asked to take part in an experiment which, it was hoped, would yield information on some new tests and they were told that their presence on both occasions was necessary if their participation was to be of value.

At the second session, the test room was divided into four sections, each subject being directed to his own section as he entered the room. All subgroups were started on their

first test at the same time and were asked to stop working on their second test at the same time. After 10 min., the A 1's had their papers removed and were presented with A 2 tests, which they began as soon as they had written their name on the cover; the B 1's were similarly changed to B 2's after 15 min; and so on. The groups were told at the beginning how long they would have for their first test but were given no information about the time limit on their second test. The complete schedule is shown in the Appendix, Table I.

It was hoped that by designing the experiment in this way, data would be gained on both immediate and successive adaptation to level of difficulty. If the hypothesis of *immediate* adaptation were confirmed, then the AH 5 problems common to an easy and a difficult test would tend, on first AH 45 testing, to be answered more correctly in the difficult than in the easy context. If *successive* adaptation occurred, that is, between two tests of different levels taken consecutively, then subgroups passing from an easy to a harder test would produce relatively low scores on comparable questions in their second test; whereas subjects beginning with their more difficult test would be likely to score relatively highly on the crucial questions in their second test.

The prediction with regard to AH 4 questions was less clear. These questions were included as creators of an easy context rather than for their own intrinsic interest. It was thought possible, however, that the response to them too might be affected by *their* context; that even these simple problems might yield results determined in part by the difficulty of the problems preceding and succeeding them.

Equated rather than identical groups were used, in order to obviate the confusion of practice effects with adaptation effects. As will be seen, the results on second testing did indicate effects on *speed* of work, which seemed to be due to a combination of practice and of differing time limits.

IV. THE SUBJECTS

The experiments reported in this paper were conducted in the U.S.A. A pilot experiment on similar lines, carried out by the writer before her visit to America, suggested that the hypothesized adaptation did take place and that it might occur more markedly in highly intelligent than in less intelligent groups (at least with the high-grade tests which she was using). She was anxious therefore to conduct inquiries with American groups of differing intellectual levels.

Three groups were chosen: graduates of research status at a private university in the West; undergraduates, mostly in their second year, at the same university; and men and women 'basic entrants' to an Air Force base in the South. The graduates were all volunteers; the majority were scientists but they included also historians, educationists and mathematicians. Their average age was just under 29 years. Of the 61, 47 were men; the 14 women were distributed as evenly as possible among the four subgroups.

119 undergraduates took part in the inquiry. Many of these were majoring in Psychology and these participated largely in order to make up their requisite number of 'experimental' hours. A substantial minority, however, were volunteers majoring in a wide variety of subjects. The average age of this group was 19 years. Here the women preponderated over the men, 71 to 48.

The 114 American Air Force subjects comprised 58 male and 56 female 'basic entrants'. These were not volunteers: the tests were included as part of their first fortnight's training.

which consisted mainly of lectures and of tests, some 'operational' and some (such as those reported on here) 'experimental'. The average age of this group was twenty years.

The testing conditions for these subjects were far less satisfactory than were those of the university groups. For example, the Air Force test groups were bigger and the supervision less adequate; many of the subjects were obviously lacking in sleep (it transpired later that some had been up since 4.00 and even 3.00 a.m.) and, understandably, were lacking in interest and incentive.

Those university subjects who were interested had the purpose of the inquiry explained to them after they had all completed their second session. There was no opportunity for such explanation with the Air Force subjects. The writer merely requested their help (which they knew they were not entitled to refuse) in 'trying out a new series of tests' and she did not have the opportunity of seeing them again after the data had been obtained.

V. RESULTS

Table 2 shows the mean scores gained by the university subgroups and the Air Force subgroups on their Pattern Perception test, first AH 45 and second AH 45 testing. It may be seen that the Pattern Perception means and standard deviations for both sets of subgroups are sufficiently similar to justify treatment of these subgroups as identical. None of these intra-subgroup differences approaches statistical significance. The figures for the graduates and undergraduates in separate groups are shown in the Appendix, Table II. Since the differences in test score between the graduates and the undergraduates were so very slight, it was thought best to amalgamate them into one large 'university group'. It may be seen that combining the two groups had the further advantage of rendering still more homogeneous the standard deviations of the subgroups.

Table 2. *Pattern Perception and AH 45 test results*

n	Pattern perception		AH 45			
			First testing		Second testing	
	Mean	S.D.	Test	Mean	Test	Mean
University						
48	48.96	6.60	A 1	30.98	A 2	24.23
45	49.69	6.88	B 1	28.98	B 2	27.09
43	49.67	6.18	B 2	27.00	B 1	29.30
44	49.59	7.28	A 2	21.95	A 1	25.32
Air Force						
30	38.23	8.09	A 1	20.63	A 2	16.00
26	37.58	8.72	B 1	20.04	B 2	19.69
28	38.07	8.32	B 2	17.21	B 1	18.61
30	37.90	7.53	A 2	16.60	A 1	17.67

If we examine the mean AH 45 scores on first testing, we find the following order of difficulty for the two main groups: A 1 (easiest), B 1, B 2, A 2 (hardest). This is the objectively correct order of difficulty for the four tests, judged by the respective numbers of AH 5 questions. Evidently the more generous time limits allowed for tests A 2 and B 2 do not compensate for the higher proportions of harder questions which they contain. The trend would presumably have been the same, with a steeper slope, had a uniform time limit been imposed for the four tests.

In Table 3 are shown the mean numbers of questions attempted by the two main groups, on first and second testing, respectively. An 'attempted' question is one which has been answered, correctly or incorrectly, or which has been omitted, at least one later question having been answered. An omitted question was classed as 'attempted' since the subject had evidently considered it at least long enough to decide to refuse it before going on to the next item. The subjects were asked to work through the problems in order, and most of them appeared to follow this instruction. Thus the figures in Table 3 probably reflect fairly accurately differences in rate of work.

Table 3. *Mean number of questions attempted*

Test	Time limit (min.)	University group		Air Force group	
		First testing	Second testing	First testing	Second testing
A 1	10	35.7	28.3	28.4	23.5
B 1	15	36.7	35.9	32.3	32.6
B 2	20	36.2	37.1	33.6	35.6
A 2	25	32.5	36.1	36.8	37.8

The interesting point about Table 3 is the influence that the time limit of the first testing in every case exerted on the subject's speed in his second testing (see arrows). The short time limits of A 1 and B 1 produced an *increased* rate of work in the second session; the relatively long time limits of B 2 and A 2 produced a *slower* rate of work in the second session. It may be seen that this effect was quite consistent, though it was more marked for some groups and some pairs of tests than for others.

This result may be seen more clearly if instead of mentally following the *subgroups* from their first to their second test, we consider the *tests*, taken first by one subgroup and then by another, equated, subgroup. Thus we see that A 1 in the university group, for instance, is tackled far more slowly (mean number attempted = 28.3) when it is taken after A 2 than when it is taken first (mean number attempted = 35.7); whereas A 2 is tackled more quickly when it is taken second (mean number attempted = 36.1) than when taken first (mean number attempted = 32.5). This result holds in every case except the Air Force group on B 1.

The writer considers this finding important for two reasons. First, it illustrates once again (Heim, 1954*a, b*) the artificial nature of designating particular tests as speed tests and particular groups as fast or slow, *in vacuo*, regardless of instructions and of the total test situation. Secondly, it provides some evidence for the somewhat complicated interpretation given below of the results obtained from the second testings in this inquiry.

It may be seen in Table 2 that on second, unlike on first, testing the means of the four tests follow no obvious pattern. The orders of difficulty for the Air Force and the university groups show no particular trend nor do they agree. They have in common only A 2 as the lowest mean, on second testing (as on first).

The combined data in Tables 3-5, and in the completed answer sheets, suggest that the lack of definite pattern in the second testing scores may be due to a conflict between two tendencies. On the one hand, there is some practice effect from the immediately preceding first testing, as described above: this results primarily in a slowing down (with consequent decreased opportunity for scoring) of those subjects who took first a harder test, with it

longer time limit; and a speeding up (with increased opportunity for scoring) of those who took first their easier test, with its shorter time limit. On the other hand, there is some successive adaptation to level of difficulty: this results in augmented power, or accuracy, *per question*, on the part of those subjects who took first their harder test and diminished power, or accuracy, *per question*, for those who took first their easier test.

This successive adaptation in turn conflicts with immediate adaptation. These latter points emerge only on examination of the mean scores per question, as shown in Tables 4 and 5. The mean in each case has been calculated by dividing the number of correct answers to each question by the number of subjects who attempted the question—not by the total number of subjects in the group. This method was found necessary in order to allow for the considerable variation in the number of questions attempted among the members of each group. The figures showing the means based on total number of subjects per subgroup are given in the Appendix, Tables III and IV.

Table 4. *Mean scores on classified test questions*¹

n	Test	AH 45 questions			AH 5 questions		AH 4 questions	
		No. of questions	Mean	No. of questions	Mean	No. of questions	Mean	
University								
48	First testing	A 1	40	0.87	6	0.50	34	0.94
45		B 1	40	0.79	14	0.51	26	0.94
43		B 2	40	0.75	22	0.58	18	0.96
44		A 2	40	0.68	30	0.59	10	0.94
44	Second testing	A 1	40	0.89	6	0.58	34	0.95
43		B 1	40	0.82	14	0.59	26	0.94
45		B 2	40	0.73	22	0.55	18	0.95
48		A 2	40	0.67	30	0.58	10	0.95
Air Force								
30	First testing	A 1	40	0.73	6	0.35	34	0.79
26		B 1	40	0.62	14	0.37	26	0.76
28		B 2	40	0.51	22	0.31	18	0.76
30		A 2	40	0.45	30	0.32	10	0.85
30	Second testing	A 1	40	0.75	6	0.30	34	0.83
28		B 1	40	0.57	14	0.28	26	0.74
26		B 2	40	0.55	22	0.34	18	0.82
30		A 2	40	0.42	30	0.31	10	0.77

¹ Means based on number of questions attempted

Let us first consider immediate adaptation to level of difficulty. This entails inspecting the results on first testing, shown in Table 4. The order of difficulty of the four tests, as shown by the mean score per question gained by each subgroup, is, as mentioned above, evidently determined by the number of AH 5 items included in each of the tests (yielding the order A 1, B 1, B 2, A 2). When, however, we turn to the corresponding mean scores for AH 5 questions alone on first testing, we find that for the university group this order is reversed. It is now as follows: A 2 (highest), B 2, B 1 and A 1 (lowest mean score).

In order to establish the relationship between the mean AH 5 score and the number of AH 5 items present in a test, the linear regression of the former on the latter was estimated. This was found to have a slope significantly different from 0: $P=0.05$ using a one-tailed test.

It is suggested that this result is due to the adaptation of subjects to the level of difficulty of the test with which they are presented. For example, subjects taking A 1 as their

first test have only six AH 5 questions among their forty: they therefore gain a higher overall score (0.87) than do members of the other three subgroups. On the other hand, they become attuned or adapted, as they work through the test, to a lower level of difficulty and therefore, despite their highest overall score, they obtain the lowest score of the four groups on AH 5 items (0.50). The position is exactly reversed for the subjects taking A 2 first with its high proportion of AH 5 items (thirty out of forty): this group gains the lowest overall score (0.68) and yet achieves the highest AH 5 score (0.59).

It will be remembered (Table 1) that every question appeared in two tests, once in A 1 or A 2 and again in B 1 or B 2. Let us consider the AH 5 questions from this point of view. On the hypothesis of upward adaptation, it should be found that when identical AH 5 questions are presented (to equated groups) in two tests, whose other items differ in level of difficulty, the AH 5 questions tend to be answered more correctly when they appear in the harder context than when they appear in the easier context.

Table 5. Mean scores on questions common to two tests¹

AH 5 questions in common					AH 4 questions in common				
No. of questions	Mean				No. of questions	Mean			
	A 1	B 1	B 2	A 2		A 1	B 1	B 2	A 2
University									
First testing	—	—	—	—	16	0.94	0.94	—	—
	6	0.50	—	0.66	18	0.93	—	0.96	—
	14	—	0.51	—	10	—	0.95	—	0.94
	16	—	—	0.55	0.55	—	—	—	—
Second testing	—	—	—	—	16	0.94	0.93	—	—
	6	0.58	—	0.58	18	0.96	—	0.95	—
	14	—	0.59	—	10	—	0.96	—	0.95
	16	—	—	0.54	0.56	—	—	—	—
Air Force									
First testing	—	—	—	—	16	0.76	0.74	—	—
	6	0.35	—	0.34	18	0.82	—	0.76	—
	14	—	0.37	—	10	—	0.79	—	0.85
	16	—	—	0.30	0.30	—	—	—	—
Second testing	—	—	—	—	16	0.82	0.73	—	—
	6	0.30	—	0.40	18	0.84	—	0.82	—
	14	—	0.28	—	10	—	0.74	—	0.77
	16	—	—	0.32	0.32	—	—	—	—

¹ Means based on number of questions attempted.

This does in fact occur with the university group (see Table 5), on first testing. Examination of the three pairs of tests possessing AH 5 questions in common, reveals that the harder test of the pair produces a higher mean score on those questions than the easier test of the pair, in all cases but one. In order to test the significance of the differences among the member-questions of the pairs of AH 5 means, a count was made of the number of times the increase was in the predicted direction. On the null hypothesis this should be about half of them; it was therefore possible to use a (one-tailed) binomial test of significance. In the case of A 1 (easiest test) and B 2 (second hardest), six of the six questions showed the predicted increase ($P=0.016$); in the case of B 1 (second easiest) and A 2 (hardest test), twelve of the fourteen questions showed it ($P=0.007$).

The pair of tests which does not yield this result is the B 2-A 2 pair. The question by question analysis here revealed no significant difference, only seven out of the sixteen:

questions showing the predicted increase in score. It may be seen in Table 5 that the AH 5 items in common did, in fact, produce identical mean scores (0.55) in the two contexts. A 2 and B 2 are the most difficult two tests and are the only pair of tests of 'adjacent' difficulty to be compared (since A 1 and B 1 have no AH 5 questions in common). This finding is, therefore, not surprising.

The trend of university results on AH 5 questions on second testing is similar but less striking. Again two of the three pairs of tests show a difference in the predicted direction, and the tests comprising the third pair produce identical mean scores. But the differences on second testings are very slight, probably owing to the differences in speed of work induced by the differing time limits for first testing and to *successive* adaptation to level of difficulty.

The subject whose first test was difficult (i.e. A 2 or B 2) will have adapted to a high level. He will therefore embark on his second (easier) test in an 'upwardly adapted' state, doing better on the AH 5 items than his equated fellows did when taking the same easy test first. This phenomenon shows clearly if we compare the respective AH 5 mean scores gained by university subjects on first and second testing (see Table 4). We are thus comparing a test taken on one occasion as a group's first test with the same test taken on another occasion as (an equated group's) second test.

The differences are not great but, in all four pairs, they go in the direction to be expected on the hypothesis of successive adaptation. A 1 and B 1, the easier two tests, respectively produce a *higher* mean score on their AH 5 questions when taken second, that is, when the subjects are upwardly adapted, having just taken a difficult test. Whereas A 2 and B 2, the harder two tests, respectively produce a *lower* mean AH 5 score when taken second, that is when the subjects are downwardly adapted, having just taken an easy test. By means of counting the number of times this occurred in individual questions it was possible to test the significance of these findings (see Table 6).

Table 6

Test	No. of AH 5 questions	No. showing predicted <i>increase</i>	<i>P</i>
A 1	6	6	0.016
B 1	14	11	0.029
		No. showing predicted <i>decrease</i>	
B 2	22	15	0.07
A 2	30	16	—

Examination of the university AH 4 results, however, yields little of interest. These questions proved so easy for the graduates and undergraduates that the mean score per question on number attempted for every subgroup lies between 0.94 and 0.96 for both first and second testing, rendering meaningless any examination for trends or for differences between means. It can only be said that for the university group, the AH 4 items provided the requisite easy background.

The Air Force results are markedly different. Their range on AH 4 items exceeded that on AH 5 items (being 0.74–0.85 and 0.28–0.37, respectively). Interestingly, on first testing, these subjects obtain their highest AH 4 mean in test A 2, the hardest test, suggesting that AH 4 may be to the Air Force as AH 5 is to the university. But inspection of the AH 4 questions in common to pairs of tests, and of the AH 4 mean scores on second testing, reveals no interpretable pattern.

It may be seen that almost all the Air Force results on AH 5 lie between 0.3 and 0.4. Thus these subjects were answering about a third of their questions correctly. This is approaching random answering, which would yield one in five. This suggests that AH 5 is almost outside the range of most of the Air Force subjects—at least in the state of fatigue and poor motivation in which they took the test.

There are no significant trends of differences among the Air Force AH 5 scores whether we examine the score-order of the AH 5 questions on first testing or the scores gained on AH 5 questions in common. If anything, the differences on these latter questions go the 'wrong' way, the slightly higher scores being gained on the AH 5 questions in their *easier* contexts. There is thus no evidence of immediate adaptation in the Air Force results. Similarly, no sign of successive adaptation is revealed on inspecting the scores on AH 5 questions taken on first and second testing. The Air Force results reveal no consistent pattern of any kind, on either the easier or the harder test questions.

VI. DISCUSSION

A small-scale pilot experiment on the same lines, carried out with English subjects, had yielded similar results. In this case, the two groups had been respectively university students and naval ratings. In this earlier experiment, too, the more intelligent, more highly educated, volunteer subjects had adapted to level of difficulty, as defined above, whereas the less intelligent, less educated, conscripted subjects gave no evidence of such adaptation.

It is not clear whether the failure to 'adapt upward' is due mainly to the poorer motivation and/or level of intellect of the non-volunteering groups or to the *relationship* between their level and that of the tests. Previous inquiries on related problems have demonstrated the importance of this latter point (Heim, Wallace & Cane, 1949, 1950; Heim, 1954*b*). It would be interesting to repeat the experiment using AH 4 questions and still easier questions, with two groups comprising, say, naval ratings and another, duller, group. The problem of volunteers versus conscripts in this connexion also merits investigation. The writer's impression is that the naval ratings would in such a situation respond to AH 4 questions in their easy context as the university groups responded to AH 5 questions in their relatively easy context.

The evidence so far is not conclusive. The results do tally, however, with observations made both outside the psychological laboratory and in non-psychometric experimental situations (Mace, 1931; Gibbs, 1951; Mowbray, 1953; Helson, 1948). Many of the factors determining behaviour in real life situations have traditionally been ignored in the group mental testing situation. But there is no reason to assume that they do not operate; indeed, such factors may play a role in producing some of the errors and discrepancies which have been grouped together, and corrected for, under the blanket term 'unreliability'.

If it be true that some subjects unwittingly adapt upwards when faced with test problems presenting them with a serious challenge and that the same subjects adapt downwards, failing to gain 100 % when tackling far easier problems than those they have shown they can solve, then inaccuracies in the assessment of intelligence of individuals, and even of groups, may sometimes arise. Presumably this is especially likely to occur, unrecognized, when all the test problems set are of roughly the same, inappropriate, level. Such questions as the gradient of difficulty, form of test instructions and methods of scoring

weighted and unweighted) immediately present themselves as demanding of more respect than they have been accorded in the past.

It would follow that the notion of one intelligence scale along which the total population can be distributed is untenable. Such a notion implies that every individual possesses a constant amount of a static quality, measurable by standard tests of intelligence. In the writer's opinion there are other objections to this view, many of which are not easily susceptible of experimentation. But the findings of this paper, if convincing and if confirmed by other investigators, would suggest that the subject's response to the same test problems will vary; he will sometimes do himself less than justice (still on the *test* criterion) and he will sometimes do 'surprisingly well'. Experimental evidence has been adduced in favour of one possible reason for such 'test unreliability'.

Thus, a number of practical considerations would follow on acceptance of the phenomenon of adaptation to level of difficulty in mental testing. It would impose, for instance, far greater care than is usually exercised in the earlier stages of testing: before administering an intelligence test to a group, the intellectual level and homogeneity of the group, and the level of the test, would be considered in relation to one another. The aim would be to allow scope for the maximum possible upward adaptation to level of difficulty, that is, the test chosen would be as difficult for the subjects as is consistent with their understanding the problems set. This suggestion is made in the belief that it is more valuable when testing to determine the subject's upper limits than to determine—albeit unreliably—what he *cannot* do.

The phenomenon would have not only practical consequences. If proved conclusively, it might have some effect on the theory of intelligence and intelligence testing since it would throw doubt on the existence of a static entity measurable and expressible in purely quantitative terms, and isolable from other characteristics, notably those usually subsumed under 'personality'.

Limitations of space preclude a full discussion here. It seems evident, however, that if the level of difficulty of a problem with which an individual can cope is partly dependent on the level of difficulty of the context, generalizations about the subject's ability based on problem-solving, without reference to the present or the immediately preceding context, are exceedingly risky; and generalizations about 'objectively based' levels of difficulty of particular problems *in vacuo* have little worth.

VII. SUMMARY AND CONCLUSIONS

1. The hypothesis was advanced that test subjects tend, unwittingly, to adapt to the level of difficulty of their test items. Upward adaptation was hypothesized when the test, whilst not grossly exceeding the limits of the subjects' ability, was sufficiently hard to extend them fully; and downward adaptation was hypothesized when the test was correspondingly simple. Previous experimenters have reported this phenomenon in non-psychometric situations.

2. In an attempt to verify this hypothesis in the field of mental testing, two pairs of tests were prepared using a selection of items from one high-grade and one easier intelligence test. The four tests comprised the same number of questions but they differed in their number of 'difficult' problems and in their time limits, the harder tests being allotted proportionately longer time limits. A group of subjects was selected, given a

preliminary, standard test of intelligence and, on the basis of their performance, divided into four equated subgroups. Each of these subgroups took one or other pair of 'mixed' tests in one or other order. Thus identical problems were presented, in contexts of varying degrees of difficulty, to equated subjects.

3. The experiments were carried out in the U.S.A., the two main groups who provided the data consisting of: *A*, graduates and undergraduates at a private university in the West; *B*, basic entrants at a U.S. Air Force base in the South.

4. The main results yielded by group *A* were as follows:

(i) Evidence of *immediate* adaptation to level of difficulty was found when a comparison was made of the scores gained on difficult questions in an easy context and a difficult context, respectively. The mean score on the difficult items was found to bear a statistically significant relationship to the proportion of difficult to easy items in the test: the more numerous the harder questions, the higher their mean score (p. 217). Moreover, identical difficult problems were solved more often (by equated subjects) when they appeared in their harder context (p. 218).

(ii) The data suggested also the operation of *successive* adaptation to level of difficulty. Subjects taking first the easier test of the pair gave evidence of some downward adaptation, achieving fewer correct solutions to the difficult problems in their (harder) second session than did members of the equated group when taking this (hard) test for the first time. Similarly, upward adaptation was demonstrated by those who took first their more difficult test: these subjects showed greater accuracy per question on the difficult problems in their (easier) second session than did equated subjects taking this (easy) test first (p. 219).

(iii) The time limit imposed on the test taken first appeared to influence the speed with which subjects tackled their second test. Those whose first test was easy, and therefore short, worked faster, attempting a greater number of items, in their second test (although it was more difficult); whereas those whose first test was hard, and therefore long, worked more slowly on their second test (although it was easier) (p. 216).

5. The data yielded by group *B*, the Air Force subjects, revealed no interpretable pattern. Their responses to the difficult questions approached sufficiently closely to random answering to suggest that with subjects of such intellectual level, tests of a lower standard should be used.

6. A small-scale pilot experiment previously conducted in England with the same two pairs of tests on a group of university students and a group of naval ratings, had produced substantially similar results.

7. It is suggested that the negative findings of the duller groups are due to the harder test problems being virtually outside their range, and that it would be worth while repeating the investigation with such groups using lower grade tests.

8. Some of the implications of the phenomenon of adaptation to level of difficulty in mental testing are briefly discussed.

VIII. ACKNOWLEDGEMENTS

The writer's thanks are due to the administrators of the Smith-Mundt and Fulbright Grants and to the Medical Research Council for financial support; to the Director of the Human Resources Research Centre at the Air Force base and of Stanford University

Psychology Department for making available the test groups and testing facilities; to the subjects for providing the data; to Prof. L. S. Penrose for allowing the use of his Pattern Perception Test; to Mr J. W. Whitfield for his help in designing the experiment; and to Miss Violet Cane, Mr E. G. Chambers and Miss Kathleen Watts for their criticism and their advice on the statistical analysis of the results.

APPENDIX

Table I. *Schedule of testing*

Subgroup	1st testing immediately followed by 2nd testing		Total time
i	A 1 (10 min.)	A 2 (25 min.)	35 min.
ii	B 1 (15 min.)	B 2 (20 min.)	35 min.
iii	B 2 (20 min.)	B 1 (15 min.)	35 min.
iv	A 2 (25 min.)	A 1 (10 min.)	35 min.

Table II. *Pattern Perception and AH 45 test results*

n	Pattern perception		AH 45			
			First testing		Second testing	
	Mean	S.D.	Test	Mean	Test	Mean
Graduates						
16	49.75	7.19	A 1	33.62	A 2	29.06
15	49.27	7.93	B 1	27.40	B 2	28.20
15	49.75	8.12	B 2	27.00	B 1	28.20
15	49.60	7.36	A 2	24.60	A 1	26.33
Undergraduates						
32	48.56	6.36	A 1	29.66	A 2	21.81
30	49.90	6.43	B 1	29.77	B 2	26.53
28	49.75	5.01	B 2	27.00	B 1	29.89
29	49.59	7.37	A 2	20.59	A 1	24.79

Table III. *Mean scores on classified test questions*¹

n	Test	AH 45 questions			AH 5 questions		AH 4 questions	
		No. of questions	Mean	No. of questions	Mean	No. of questions	Mean	
University								
48	First testing	A 1	40	0.77	6	0.45	34	0.83
45		B 1	40	0.72	14	0.48	26	0.86
43		B 2	40	0.68	22	0.53	18	0.85
44		A 2	40	0.55	30	0.48	10	0.75
44	Second testing	A 1	40	0.63	6	0.41	34	0.67
43		B 1	40	0.73	14	0.54	26	0.84
45		B 2	40	0.68	22	0.52	18	0.87
48		A 2	40	0.61	30	0.52	10	0.85
Air Force								
30	First testing	A 1	40	0.52	6	0.26	34	0.56
26		B 1	40	0.50	14	0.30	26	0.61
28		B 2	40	0.43	22	0.27	18	0.63
30		A 2	40	0.42	30	0.30	10	0.77
30	Second testing	A 1	40	0.44	6	0.18	34	0.49
28		B 1	40	0.46	14	0.23	26	0.59
26		B 2	40	0.49	22	0.30	18	0.73
30		A 2	40	0.40	30	0.29	10	0.72

¹ Means based on number of questions in the test.

Table IV. Mean scores on questions common to two tests¹

AH 5 questions in common						AH 4 questions in common					
		Mean						Mean			
		No. of questions	A 1	B 1	B 2	A 2	No. of questions	A 1	B 1	B 2	A 2
University											
First testing	{	—	—	—	—	—	16	0.86	0.83	—	—
		6	0.45	—	0.63	—	18	0.81	—	0.85	—
		14	—	0.48	—	0.53	10	—	0.90	—	0.75
		16	—	—	0.49	0.44	—	—	—	—	—
Second testing	{	—	—	—	—	—	16	0.68	0.80	—	—
		6	0.41	—	0.56	—	18	0.66	—	0.87	—
		14	—	0.54	—	0.55	10	—	0.89	—	0.85
		16	—	—	0.50	0.50	—	—	—	—	—
Air Force											
First testing	{	—	—	—	—	—	16	0.55	0.55	—	—
		6	0.26	—	0.29	—	18	0.57	—	0.63	—
		14	—	0.30	—	0.32	10	—	0.70	—	0.77
		16	—	—	0.26	0.28	—	—	—	—	—
Second testing	{	—	—	—	—	—	16	0.47	0.55	—	—
		6	0.18	—	0.36	—	18	0.50	—	0.73	—
		14	—	0.23	—	0.28	10	—	0.65	—	0.72
		16	—	—	0.28	0.30	—	—	—	—	—

¹ Means based on number of questions in the test.

REFERENCES

- GIBBS, C. B. (1951). Transfer of training and skill assumptions in tracking tasks. *Quart. J. Exp. Psychol.* **III**, 99-110.
- HEIM, A. W. (1947). An attempt to test high grade intelligence. *Brit. J. Psychol.* **xxxvii**, 70-81.
- HEIM, A. W., WALLACE, J. G. & CANE, V. R. (1949, 1950). The effects of repeatedly retesting the same group on the same intelligence test. Parts I, II and III. *Quart. J. Exp. Psychol.* **i**, 151-9; **ii**, 19-32 and 182-97.
- HEIM, A. W. (1954*a*). *The Appraisal of Intelligence*, chap. ix. London: Methuen.
- HEIM, A. W. (1954*b*). Les conventions à la base des tests d'intelligence. *Enfance*, 105-12.
- HEIM, A. W. (1954*c*). Psychological adaptation as a response to variations in difficulty and intensity. Unpublished.
- HELSON, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.* **lv**, 297-313.
- MACE, C. A. (1931). The influence of indirect incentives upon the accuracy of skilled movements. *Brit. J. Psychol.* **xxii**, 101-14.
- MOWBRAY, G. H. (1953). Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty. *J. Exp. Psychol.* **xlvi**, 365-72.
- PENROSE, L. S. (1944). An economical method of presenting matrix intelligence tests. *Brit. J. Med. Psychol.* **xx**, 144-6.
- PENROSE, L. S. (1946). The Pattern Perception Test: experimental results. *Brit. J. Med. Psychol.* **xx**, 339-43.
- WALLACE, J. G. (1952). The results of a high grade intelligence test applied to a university population. *Brit. J. Psychol.* **xlili**, 61-9.
- WATTS, K. P. (1953). Influences affecting the results of a test of high grade intelligence. *Brit. J. Psychol.* **xliv**, 359-67.

(Manuscript received 30 July 1954)

A TEST OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEANS OF CORRELATED MEASURES, BASED ON A SIMPLIFICATION OF STUDENT'S t

BY JOSEPH SANDLER

The Tavistock Clinic, London

Psychologists have often to compare the means of two distinct sets of scores, where there is a correlation between the two sets. This is usually done when:

- (i) Two groups have been matched with respect to some extraneous variable, such as age or intelligence. Each person in one group is paired with a person in the other.
- (ii) A set of persons has been given a test twice, in different circumstances (e.g. before and after training, or the administration of a drug).

The appropriate test of significance of the difference between the two means is Student's t by the method of differences (or by the more laborious method in which the correlation between the two sets of scores is actually calculated and then corrected for). t by the method of differences is given as follows (Johnson, 1949):

$$t = \frac{\bar{d}}{\text{S.E.}_{\bar{d}}},$$

where $\bar{d}$ is the mean difference between the pairs of scores, and $\text{S.E.}_{\bar{d}}$ is the standard error of the mean difference. t has $N - 1$ degrees of freedom, where N is the number of cases in each group.

It is the purpose of this paper to present a statistic A , which yields exactly the same results as Student's t , being rigorously derived from it, but which is extremely simple to calculate. If, as is commonly the case, two matched groups are compared with respect to a large number of variables, the use of A results in a considerable saving of time and labour. A is found as follows:

$$A = \frac{\text{The sum of the squares of the differences}}{\text{The square of the sum of the differences}} = \frac{\sum d^2}{(\sum d)^2}.$$

A table of A has been constructed, and is appended. The table is entered with $N - 1$ degrees of freedom, and a value of A is significant at any given level if it is equal to or less than the tabled value. It corresponds to the two-tail form of t .

If we write t as

$$t = \frac{(\sum d)/N}{\sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{N}}{N(N-1)}}},$$

taking the square of both sides gives us

$$t^2 = \frac{N-1}{N} \frac{(\sum d)^2}{\sum d^2 - \frac{(\sum d)^2}{N}}.$$

Hence,
$$\frac{N-1}{Nt^2} = \frac{\Sigma d^2}{(\Sigma d)^2} - \frac{1}{N}.$$

From this we may write A in terms of t :

$$A = \frac{N-1}{Nt^2} + \frac{1}{N},$$

t can be written in terms of A as follows:

$$t = \sqrt{\frac{N-1}{AN-1}}.$$

The table of A has been constructed from published tables of Student's t (Fisher & Yates, 1948) and the tables of the inverted beta (F) distribution, computed by Merrington & Thompson (1943).

Table of A

(For any given value of $N-1$, the table shows the values of A corresponding to various levels of probability. A is significant at a given level if it is equal to or less than the value shown in the table.)

N-1	Probability					N-1
	0.10	0.05	0.02	0.01	0.001	
1	0.5125	0.5031	0.50049	0.50012	0.5000012	1
2	0.412	0.369	0.347	0.340	0.334	2
3	0.385	0.324	0.286	0.272	0.254	3
4	0.376	0.304	0.257	0.238	0.211	4
5	0.372	0.293	0.240	0.218	0.184	5
6	0.370	0.286	0.230	0.205	0.167	6
7	0.369	0.281	0.222	0.196	0.155	7
8	0.368	0.278	0.217	0.190	0.146	8
9	0.368	0.276	0.213	0.185	0.139	9
10	0.368	0.274	0.210	0.181	0.134	10
11	0.368	0.273	0.207	0.178	0.130	11
12	0.368	0.271	0.205	0.176	0.126	12
13	0.368	0.270	0.204	0.174	0.124	13
14	0.368	0.270	0.202	0.172	0.121	14
15	0.368	0.269	0.201	0.170	0.119	15
16	0.368	0.268	0.200	0.169	0.117	16
17	0.368	0.268	0.199	0.168	0.116	17
18	0.368	0.267	0.198	0.167	0.114	18
19	0.368	0.267	0.197	0.166	0.113	19
20	0.368	0.266	0.197	0.165	0.112	20
21	0.368	0.266	0.196	0.165	0.111	21
22	0.368	0.266	0.196	0.164	0.110	22
23	0.368	0.266	0.195	0.163	0.109	23
24	0.368	0.265	0.195	0.163	0.108	24
25	0.368	0.265	0.194	0.162	0.108	25
26	0.368	0.265	0.194	0.162	0.107	26
27	0.368	0.265	0.193	0.161	0.107	27
28	0.368	0.265	0.193	0.161	0.106	28
29	0.368	0.264	0.193	0.161	0.106	29
30	0.368	0.264	0.193	0.160	0.105	30
40	0.368	0.263	0.191	0.158	0.102	40
60	0.369	0.262	0.189	0.155	0.099	60
120	0.369	0.261	0.187	0.153	0.095	120
∞	0.370	0.260	0.185	0.151	0.092	∞

REFERENCES

FISHER, R. A. & YATES, F. (1948). *Statistical Tables*. London: Oliver and Boyd.
JOHNSON, PALMER O. (1949). *Statistical Methods in Research*. New York: Prentice Hall.
MERRINGTON, MAXINE & THOMPSON, CATHERINE M. (1943). Tables of the inverted beta (F) distribution. *Biometrika*, xxxiii, 73-88.

(Manuscript received 10 February 1955)

CORRESPONDENCE

THE EDITOR

British Journal of Psychology

Dear Sir,

In his recent paper,¹ Mr A. S. C. Ehrenberg says: 'Scott Blair (1949)² adopts a common point of view when he says that it is clear that three dimensions are required to describe space. However, this is not only not clear, but demonstrably false.'³ Considering, as a two-dimensional example, my description of a symmetrical American city built on flat ground, with streets at right-angles, he says: 'one-dimensional descriptions could be made up, for example, as follows:

1	2	3	4	5	6	7	8
9	10	11	12	.	.	.	.
17	18	19	.	.	.	.	.

With this latter map, one could find one's way around....'

However, he has not, in this way succeeded in describing position in a plane by means of a single dimension. All he has done is to 'hide' one of the dimensions. To find my way to no. 18, for example, I require not only the number 18, but also the number 8, which tells me where to start looking for any number larger than 8. Of course, if the whole city consists of a convenient square of 8 blocks, this may be obvious; but, 'to describe space', one should use a method which is not limited, in this way, to finite distances of arbitrary magnitude.

In American hotels, all the rooms on, say, the 20th floor are often numbered upwards from 2000 (2001, 2002, ..., etc.), even though there may be considerably less than 100 bedrooms per floor. Then no room is numbered 2099 and the first room on the 21st floor will be 2100 or, better, 2101. This is another convenient version of Mr Ehrenberg's system, or something rather like it; but, of course, each of these numbers is really complex. In the case of, say, 2143, the 21 represents directly the vertical dimension and the 43, a mixture of the two horizontal dimensions. Even on the 21st floor, we cannot tell where the room 43 will be without some other 'hidden' number which, like the '8' in Mr Ehrenberg's example, tells us how many rooms there are along each corridor, or gives us some similar architectural information. Thus no. 18 in Mr Ehrenberg's city should really be number 8:(8) and my room in the hotel 21:43:(45). The figures in brackets would, of course, require further explanation. In the latter case, (45) might tell me that my room was the third from the far end of the corridor.

The point is that we still have our two magnitudes to describe position in a plane and three to locate a point involving also altitude. (I do not, of course, imply that highly accurate measurements made over exceptional distances may not require the use of geometries more complex than that of Euclid.)

¹ Ehrenberg, A. S. C. (1955). Measurements and Mathematics in Psychology. *Brit. J. Psychol.* XLVI, 20-9.

² Scott Blair, G. W. (1949). *Measurements of Mind and Matter*. London: Dobson.

³ Mr Ehrenberg was kind enough to make this criticism to me some time ago in a private conversation, but could not immediately frame an adequate reply and I was most unfortunately not able to be present when he read his paper in Nottingham.

I willingly admit that my American city analogy, like all analogies, is imperfect. First, even the largest American cities have arbitrary boundaries and, as I have already said, we should not assume arbitrary boundaries in selecting fundamental methods for measuring space. Secondly, one might be tempted to think only in terms of discrete units—blocks, houses, etc., and to forget that, unless we are dealing with quantum physics, there should be no limit to the precision of our measurements. This point is covered by adding any necessary number of decimal places to our figures. Thus, instead of

1,1	1,2	1,3	.	.	.
2,1	2,2	2,3	.	.	.
.	.	.	.	.	.

we can write

1.0, 1.0	1.0, 1.1	1.0, 1.2	.	.
1.1, 1.0	1.1, 1.1	1.1, 1.2	.	.
.	.	.	.	.

If we could be content with discrete units, we might use a method somewhat like that of the 'arrondissements' in Paris.

7	8	9	
6	1	2	etc.
5	4	3	

We can make our units as small as we please by using enough numbers to cover any given area, but, whatever the size of the unit, the 'key' to such a method involves stipulating a 'quantum': the smallest area which can be located and, if we alter the size of the quantum, we alter the whole system.

We may be tempted to take a fairly large unit and to divide this into, say, nine parts, any number of times, allocating second, third and subsequent digits in a similar spiral fashion. But the selection of nine is itself as arbitrary as is the size of the original units and the method will not work unless some such number is chosen.

This is the nearest I can get towards describing location in a continuum plane by a single number. Perhaps Mr Ehrenberg can do better; but the matrix proposed in his paper certainly does not help.

Whether it is false to say that 'three dimensions are required to describe space' thus depends rather (as the late Prof. Joad would have said) on what you mean by 'describing space'.

I should add that with much of the rest of Mr Ehrenberg's argument I am in the heartiest agreement. Too rigid a theory of dimensions can have a crippling effect on the development of physics and, in some respects has already done so. Indeed my main purpose in writing my little book was to warn psychologists and others of this danger.

Yours faithfully,

G. W. SCOTT BLAIR

Physics Department

National Institute for Research in Dairying

University of Reading

17 March 1955

THE EDITOR,
British Journal of Psychology

Dear Sir,

In his stimulating letter, Dr G. W. Scott Blair suggests that the point 18 in my illustrative one-dimensional space-description might be called 18 (8). By this he seeks to demonstrate the existence of additional dimensions. But he is in fact developing rules and laws telling us how to get from one point to another, and in doing so has used only the values of the one dimensional scale, which therefore remains quite sufficient. Admittedly there are implicit in this one-dimensional system of measurement other, multi-dimensional ones. But then any one measuring system necessarily 'hides' all others that are possible. My own concern of course lay with the given measuring system, in its own right, overtly.

Dr Scott Blair's claim that 'we still have our two magnitudes to describe position in a plane and three to locate a point involving also altitude' would only be true—and then it would amount to a tautology—if the concepts of a certain kind of 'plane' (e.g. a horizontal one) and of 'altitude' were to be immediately explicit in the measuring system as defined: this dimension means altitude, those two define a horizontal plane. But with other measuring systems such concepts could make their appearance later, in the all-important development phase. This Dr Scott Blair has himself indirectly illustrated, and more generally, concepts like *inclined* planes, for example, would only emerge later even for the three-dimensional system that he has in mind.

The finer subdivision of a domain of measurement, and its possibly infinite extent—points which Dr Scott Blair has also touched upon—do not, I think, raise different issues for a one-dimensional space-description from those for a three-dimensional one. The theorem referred to in the paper, that all rational points are enumerable, seems to cover it all and since Cantor's original proof amounted to showing that rules for such enumeration exist, any attempt at their further elaboration would be an academic exercise which I myself would like to avoid.

Yours faithfully,

A. S. C. EHRENBERG

EDITORIAL NOTE

We have received a letter from Dr D. J. Finney, University of Aberdeen, commenting on the reference made to him by C. C. Anderson in his paper 'Some simple methods of testing for function fluctuation' (this *Journal*, XLVI, part 1, 1-12). Dr Finney points out that the 'memorandum' quoted was in fact a note prepared for the Director of the National Foundation for Educational Research in England and Wales (at his request) in comment on a confidential report submitted by Mr Anderson, and was intended solely as a basis for private discussion between the Director, Mr Anderson and himself. Dr Finney strongly deprecates any suggestion that his note represented a distinctive theory of function fluctuation and asks us to state that he would never have consented to a published discussion of a hurriedly prepared private document on a subject of which he lacked expert knowledge.

Mr B. S. Morris, Director of the National Foundation, has indicated that he is in entire agreement with Dr Finney's remarks.

Mr Anderson writes: 'I regret very much that my paper has given offence to Dr Finney. I was entirely unaware that his document was confidential and preparatory. If I have misrepresented his views I offer my sincere apologies.'

Mr Anderson has also written to correct the formula which appears on p. 8 of his paper. This should read:

$$\frac{\frac{B-C}{p}}{\frac{B-C}{p} + C} = \frac{\frac{1}{p} \left(\frac{B}{C} - 1 \right)}{\frac{1}{p} \left(\frac{B}{C} - 1 \right) + 1} = \frac{I_2}{I_2 + 1}.$$

PUBLICATIONS RECENTLY RECEIVED

Probleme der ganzheitspsychologischen Wahrnehmungslehre. By WALTER EHRENSTEIN. Leipzig: Johann Ambrosius Barth. 1954. Pp. ix + 342. Paper, DM. 19.50; bound, DM. 21.40.

This book, by one of the best known followers of the Gestalt school, is a new (third) edition, considerably changed from that which appeared in 1942. To experimental workers in the perceptual field it will be a store-house of brilliantly conceived experiments, interesting discussion, and somewhat biased theoretical speculation. It shares with its predecessor the same virtues and vices. Modern American literature is almost completely disregarded; theoretical discussions are twenty to thirty years out of date. The book certainly cannot be used, as the author claims, as a text-book; students would get completely erroneous views of modern theories in perception. Used rather as a compendium of interesting and often important phenomenological observations, this paper-backed volume is well worth its price. However, the reader would almost certainly be required to re-interpret the facts in terms of more modern and more consistent principles than those contributed by Ehrenstein.

The book may be of interest also to workers in the field of personality as the author has used his background in the perceptual field as a basis for a well-known typology. There has been considerable interest in recent years in linking up these two fields, and while the methodology used by Ehrenstein for the purpose of establishing his typology (cf. his book entitled *Grundlegung einer ganzheitspsychologischen Typenlehre*) is not adequate for the purpose, nevertheless students will find many hypotheses emerging from Ehrenstein's discussion which are worthy of being put to an experimental test.

H. J. EYSENCK

Physical and Psychical Research: an Analysis of Belief. By C. C. L. GREGORY and ANITA KOHSEN. Reigate: Omega Press. 1954. Pp. ix + 213. 15s.

The main difficulty about psychical research is that its results indicate the occurrence of events which are contrary to the expectations based on the acceptance of an orthodox scientific system of belief. Two reactions are possible to this situation. One is to adhere to the belief system and to reject the alleged facts as impossibilities. The other is to question the foundations of the system of beliefs which has led one to make the judgement which leads to rejection.

Every increase in the strength of evidence for at least some paranormal phenomena is an additional reason for preferring the second alternative, and the strength of evidence has increased during the last thirty years to an extent not always realized. The authors of this book (a physicist and a psychologist) accept the necessity for the second of these alternatives, and they make a searching examination of the orthodox science belief system. Their position is summed up on the last page but one: 'Science as a discipline is magnificent and indispensable; as a belief system it is disastrous.' They contend that the implicit belief system of the Western World exerts a general defacilitation of images other than those related to material systems.

The authors envisage the possibility of a unified science which will include the facts of parapsychology and serve to bridge the gap between rival systems of belief. As a means towards this end, they suggest the necessity for the development of 'metascience' which would have as its subject-matter the informational relations between centres, individuals, and groups. The present book may be regarded as preliminary spade-work in preparation for these ends. It is to be recommended to those who are interested in the revaluation of their own belief systems.

E. H. THOULESS

Psychological Statistics. By QUINN MCNEMAR. New York: Wiley; London: Chapman and Hall. Second edition. 1955. Pp. vi + 408. 48s. net.

The first edition was reviewed in this *Journal* (1950, xli, 87-8) by Patrick Slater. The author had not originally intended the book for use in introductory courses, and changes in the new edition are designed to make it more suitable for this purpose. Nevertheless, to the present reviewer Mr Slater's earlier criticism still appears to be valid: that the presentation is a little difficult for the mathematically unsophisticated.

In addition, some topics (e.g. Discriminant Function, pp. 210–11) are treated so concisely that one is left uncertain both of what the technique is designed to accomplish and of how the actual computation is carried out.

Throughout, the author's emphasis is on testing of hypotheses as the primary aim in the application of statistical inference, with the binomial distribution as the preferred starting-point of his exposition. The development of this theme has led the author to rewrite the chapters on analysis of variance, which now receive much fuller treatment than in any other text of comparable length and scope. Anticipated criticism of the short space (4 pp.) allotted to currently fashionable non-parametric methods is answered by the author's statement that in his view not enough is yet known about their respective merits.

Most users of statistics have their favourite text-books; while the present reviewer is unlikely to transfer his allegiance to McNemar, he has already found the book indispensable, and the Appendix contains a particularly handy selection of Tables.

BORIS SEMEONOFF

Transformation Analysis of Factorial Data. By YRJÖ AHMAVAARA. *Annals of the Finnish Academy of Sciences*, Series B, Volume LXXXVIII, 2. Helsinki. 1954. Pp. 150. No price given.

This monograph is in three parts. In Part I the author presents his 'transformation method', of comparing different factorial studies carried out with different populations. In Part II he applies this method to a comparison of several of Thurstone's studies of mental ability. Finally, in Part III, he outlines a technique by which the residual matrix in a factor study may be analysed to detect further 'latent common factors'.

The importance of the problem dealt with in Part I cannot be denied. The author tackles it by drawing a parallel between Pearson's selection theory and Thurstone's multiple-factor theory. Then by postulating 'a theory of factorial invariance', and using an argument similar to that used by Pearson he constructs a 'comparison matrix' by which the results of the two studies can be compared. More precisely suppose there are two $n \times r$ matrices, X and Y , of factor loadings one for each of two samples, where n refers to tests and r to factors. Then if the matrix X be taken as the invariant link between the factor scores for the two samples—corresponding to the invariance of the regression plane assumed by Pearson—and the specific factors remain orthogonal to the common factors and to each other, the comparison matrix takes the form $(X'X)^{-1}X'Y$.

This form is familiar. It is the rotation matrix required to rotate the factor matrix X into maximum conformity, in the least square sense, with the matrix Y . It can be obtained quite independently of Pearson's selection theory and of the assumptions made in it. (Indeed the author seems to do so on p. 51.) The recognition of this fact is important and helps one to evaluate the author's rather far-fetched, if not misleading, statement (pp. 36–7) that 'the comparison matrix of given factors is independent of the population. This must even be the case, for what is concerned is simply that the result of comparison does not depend on the direction in which the comparison is made.' This statement is true only while the matrix X is retained as the invariant link between the two studies. This is equivalent in our terminology to saying that the matrix X has been chosen for rotation. Here an analogy may be drawn with regression for just as there are two regression lines so there are two comparison matrices, and if comparison is made in the opposite direction the new comparison matrix is $(Y'Y)^{-1}Y'X$. The author no doubt realizes this but he certainly does not make it clear. Furthermore, the analogy with regression leads one to speculate that what is required, if this approach is adopted, is something like the product of these two comparison matrices, just as the product of two regression coefficients gives the square of the correlation coefficient.

Ahmavaara, however, justifies his method by using it very effectively to test the invariance of some well-known factor results by Thurstone.

There follows a useful critical review of the simple structure concept, and of what changes are to be expected in a factor pattern with increase in age. The monograph ends with three short chapters describing how the author examines a residual matrix for latent common factors by the calculation and examination of coefficients which he calls 'residual spectra'. In this way he is able to demonstrate, again using Thurstone's data, that latent common factors that are revealed there result from the tendency of the primary factors to break up into what he terms their 'fine structure components'.

In the monograph the number of errors that occur where reference is made to earlier pages and paragraphs are too numerous to mention. One safe rule would be to read §3 for §2, and §4 for §3, almost everywhere one meets them.

All in all this monograph provides very stimulating reading and this is especially true of the non-mathematical sections. One is indebted to the author for bringing into the open at least one of the greatest limitations of the factor technique, that is the need of a satisfactory statistical test for comparing results obtained from different experiments.

A. E. MAXWELL

The Metaphysics of Logical Positivism. By GUSTAV BERGMANN. Longmans, Green and Co. 1954. Pp. x+341. 42s.

Prof. Gustav Bergmann, one of the original members of the Vienna Circle, is known among psychologists for his application of the Carnapian technique of artificial languages to psychological theory. The emphasis of this work, is however, primarily philosophical, and most of the eighteen essays collected here have been previously published in philosophical journals.

Bergmann gives a good account as well as critique of logical positivism, which explains the 'metaphysics' in the title. He points out that two distinct types of linguistic philosophy have now developed: (1) devoted to the construction of artificial languages, cultivated to-day largely in the United States; (2) the analysis of philosophical correct usage which flourishes among the therapeutic positivists in this country. He is wise enough not to regard philosophical problems as being merely verbal, but he does imply that the technique of the artificial language is the *Open Sesame* to such problems. He believes, for example, that an adequate solution of the mind-body problem exists in terms of such a logical reconstruction of experience.

His essay 'On the problem of relations in classical psychology', is of direct interest to psychologists. All the classical psychologists discussed, British writers such as Locke, Hartley, James Mill, Reid, and continentals like Wundt and Brentano, accept a common method of inquiry, introspective analysis. These writers either conceived relations as analysable into atomistic elements, or as the resultant of a creative psychological act. They however overlooked another possibility which forms the very core of the more recent Gestalt doctrine, namely the irreducibility and immediacy of relational experience. Bergmann's criticism of classical psychology is then not that it was unscientific, but that it did not adequately take account of relations as immediately given experiential elements.

Besprinkled as it is with logical formulae this book is at times far from easy reading. Nevertheless, it will repay study by those psychologists who wish to obtain some insight into the linguistic vagaries of some modern philosophers.

W. MAYS

Die Frustration in persönlichkeitspsychologischer Beleuchtung. By REINO PALAS. *Annals of the Finnish Academy of Sciences*, Series B, Volume LXXXVII, 2. Helsinki. 1954. Pp. 110. 400 Markka.

This is a general survey of the part played by the concept of 'frustration' in contemporary psychology. It is defined as 'a state of affairs in which a living being is prevented by some obstacle, external or internal, from reaching its goal'. The author goes on to consider causes of frustration, reactions to frustration, and individual differences in ways of coping with frustration. His material is drawn from Freud, from the various offshoots of Freud's work, from Lewin and his pupils, and from recent American contributions. There are also references to Scandinavian and German writings which are less accessible to ourselves.

The author's success in keeping abreast of recent developments, especially in the U.S.A., compares more than favourably with that of most continental psychologists. One must, however, admit that there are signs of his remoteness from the places where most of the work he describes is going on. To, say, an American psychologist in the thick of current controversies, he must seem occasionally to have missed the point. His monograph is likely to be valuable as a convenient means of putting German-speaking psychologists in the picture. But it will not have much that is new to tell English-speaking readers, and more satisfactory summaries of the field are readily available.

D. E. BERLYNE

Human Development. By JOHN P. ZUBEK and PATRICIA ANNE SOLBERG. New York: McGraw-Hill. 1954. Pp. vi+476. Price \$6.00.

The aim of this book, explicitly stated in the preface, is to satisfy the 'need for a text that would cover the entire age span from conception to old age, devoting about the same amount of space to each period of life'. The pattern of the book, largely determined by such an aim, emerges very clearly. Each section,

whether it deals with the senses, learning, emotion or attitudes, traces the behavioural changes which occur as the individual progresses from infancy, or earlier, through childhood, adolescence and adulthood to old age.

Approximately half of the book is physiological: the first eight chapters deal with genetic factors in behaviour, and neural, glandular, physical, motor and sensory development. For the rest, learning, intellectual development, emotional development, social development, changes of interest, attitudes and personality are given a chapter each.

The scope of the book is thus enormous, and inevitably the pint pot will not hold several gallons. Yet one would hesitate to call this a superficial book: rather, the authors have chosen to deal with each topic from one particular point of view, the developmental, and to limit themselves severely to this. There are, however, some strange omissions—a chapter on learning and symbolic processes with no reference to Hull, Mowrer or Miller; a section on intelligence and mental factors with no reference to Spearman or Burt.

The major achievement of the authors lies in bringing together, in one reasonably sized volume, a tremendous amount of material, a great deal of it the result of recent research, which previously was scattered abroad in innumerable journals and text-books. They have collated this material and presented it in a well-organized text-book, lucidly written and very well illustrated, which should prove extremely useful to students of developmental psychology. The bibliography, containing well over 750 references, is evidence of the wealth of material provided, though in a book of this kind one might argue for a separate bibliography for each chapter, rather than one covering the entire book.

ELIZABETH FRASER

A Dynamic Psychopathology of Childhood. By LAURETTA BENDER. Springfield, Illinois: Charles C. Thomas. 1954. Pp. xiii + 275. 54s.

This is the third of the Bellevue studies in Child Psychiatry. It is a collection of papers by a number of authors, Lauretta Bender and Paul Schilder being the chief contributors. The papers have in the main been previously published but are enlarged here by the addition of later follow-up material which is of considerable interest. In some cases records run from early childhood into adult life.

No systematic survey of psychopathology is attempted, but the topics dealt with are of importance in psychopathology. There is stress on the nature of the fantasy life in children, and with studies of imaginary companions, hallucinations in children and internalization of fantasy objects. Studies of children's reactions to the realities of sexual seduction and to the death of parents or siblings yield some of the more interesting information.

In general, the number of children in each study is small, less than twenty, and the case-history of each child is summarized. The reader may in the end feel that he knows far too little about far too many children, while the small numbers naturally reduce the value of the conclusions. One may wonder, for example, whether the comparatively salutary effects of pre-pubertal sexual relations found in this study is general; or again whether the breakdowns associated with failure in sex-identification would be so apparent in a population containing fewer schizophrenics. However, much of the case material is of a kind not easily to be met with and its publication is clearly of great use. Where it is given in some detail, as in the chapter on Impulsions, Compulsions and Obsessions, its value is a good deal enhanced.

ELIZABETH NORMAN

Psychosurgery and the Self. By MARY FRANCIS ROBINSON and WALTER FREEMAN. New York and London: Grune and Stratton. 1954. Pp. ix + 118. \$3.00.

This is Dr Robinson's book. Dr Freeman contributes only a very short chapter on the techniques of pre-frontal leucotomy. But he was one of the first to see that leucotomy is apt to bring about real, if elusive, changes in self awareness and to direct his psychologist to their study. And he was singularly lucky in his psychologist. To her technical qualifications, Dr Robinson adds qualities of perceptiveness and humanity as admirable as they are rare. She has had the good sense to realize that clinical testing, to have any real value, must be founded upon—indeed grow out of—intimate clinical study. As a result, her tests are based not—as is usual—upon some arbitrary consortium of faculties, traits or factorial dimensions but upon close and sympathetic study of individual cases. Her experience—and Dr Freeman's—led to the hypothesis that an important effect of leucotomy lies in reducing what may be crudely called 'self-continuity'. By this is meant that the patient no longer relates himself fully to his past or his future, lacks all subtle

appreciation of himself in relation to others, and no longer depends for his self-respect upon the acceptance or approval of his fellows. He is, in short, a simplified preparation of *Homo sapiens*.

The tests described are mainly directed to assessing in more quantitative form these changes in the leucotomized patient's attitude to himself, his past and his future. They are based largely on the extent to which the patient talks about himself (and to a lesser extent on the ways in which he talks about himself) in the course of a contrived experimental interview. Control studies are reported with comparable psychiatric patients who had improved without operation. The results undoubtedly indicate some failure of self-critique and diminished interest in one's own past and future as characteristic long-term effects of operation. The results of an additional method, quaintly called a 'Sensibility Questionnaire', point in the same direction. In general, the test results convincingly bear out the clinical observations which inspired them. They do not, however, throw any new light on the mechanisms responsible for personality change.

It will not escape the psychiatrist that the bulk of this work was carried out with patients who had undergone so-called standard leucotomy—in which the extent of surgical damage is fairly considerable. These operations are now increasingly giving place to more limited types of leucotomy, and it is notable that in the few cases reported here in which less extensive operations were carried out the disturbance in 'self-continuity' was correspondingly diminished.

This little book deserves the sympathetic attention of psychologists—and not merely of those whose routine work takes them into the clinical field. Without venturing an opinion on the rationale of psychosurgery, or on the vexed question of whether its psychiatric gains outweigh the losses, this book may be recommended to all psychologists not yet so atrophied by professional involvements as to have lost all curiosity about the human mind and its obscure dependence on the brain. Had William James been alive to-day, it is not too fanciful to suppose that he might willingly have written a Foreword.

O. L. ZANGWILL

Emotions and Bodily Changes. By FLANDERS DUNBAR. New York: Columbia University Press; London: Geoffrey Cumberlege. Fourth edition. 1954. Pp. xxii + 1192. £6.

This well-known book has now reached its fourth edition, which incorporates supplementary material and additional bibliography. The changes all appear to be in the nature of addition, without subtraction or substitution. The original theoretical introduction is still retained, although much of it has now rather an antiquated flavour.

The important new material is chiefly in the Clinical sections, under the headings of Metabolism (including problems associated with the feeding of infants and children), Genito-urinary System, Special Sense Organs (under which heading hysterical sensory abnormalities are presented), and Bones, which includes the section on Dentistry and an account of how its application in that field emptied a Dental Conference.

The main features of the book are familiar. It may justly be described as a mine of information, although not without a certain implication that some digging, sorting, and refining is necessary.

In his theoretical approach the author manages to combine the organismic and the psychoanalytical without discomfort. It is arguable, however, that the system implicit in the clinical sections is tacitly dualist.

The clinical reports, abstracted to the number of some five thousand, did not yield any controlled clinical inquiries to sampling, but this was perhaps merely the reviewer's misfortune. At all events, this type of inquiry cannot be said to bulk large in the field. No definitive answer can therefore be expected to theoretical questions.

The value of the book, then, to the academic psychologist, is largely limited to a surreptitious glance at the section on his own symptoms. To the clinical psychologist it is a five-thousandfold repetition of his own type of experience. To the practising physician, and perhaps more particularly specialist, at whom, presumably, the book is principally directed, its value could be considerable. This is, no doubt, the justification for the Clinical presentation, system by system and organ by organ, which otherwise appears somewhat clumsy and repetitive. Symptoms can be looked up as in the usual types of medical reference book, theoretical sections skipped, and suspicions crystallized.

Whether this leads to an adequate therapeutic approach is open to some doubt. This cannot, however, reasonably be regarded as a fault of the book, but is, rather, an unfortunate consequence of the spectacular triumph of the analytical approach in medicine over the past fifty years, of the short-term dividends that thinking in terms of 'diseases' rather than 'disease' has paid.

J. A. L. NAUGHTON

Medical Treatment of Mental Disease. By D. J. McCARTHY and K. M. CORRIN, with sections by eight contributors. Philadelphia, Lippincott; London: Pitman. 1955. Pp. xx + 653. 96s. net.

This book is to be welcomed, if only because the authors see mental disorders as a part of general medicine to be treated as medical problems. They see psychiatric cases as sick people to whom one should apply the fundamental rules of general medicine. For this emphasis they are to be lauded. While a number of differences in style and viewpoint of the various contributors can be detected, the orientation of the book is fundamentally in favour of the organic as opposed to the psychogenic basis of psychiatry. They assume an organic or toxic aetiology for the vast majority of mental diseases, using this for the study and treatment of the case. The concepts and treatment of mental disease are dealt with as matters of disordered chemical, physiological or pathological processes affecting brain function. Emotional factors are recognized as being important, but more often as the effects of disease rather than the causes. No attempt is made to integrate the dynamic and purely psychological aspects. The authors believe they are restricting themselves to fundamentals of proven scientific value and hold the view that only on an organic basis can psychiatry advance scientifically.

In many respects, the book is out of date: the detection of hidden infection in the body is stressed as one of the most important aspects of psychiatric diagnosis. This emphasis on focal infection will to many psychiatrists be a return to the early part of the century when ruthless eradication of 'foci' was the order of the day. The section on personality is disappointing, as is that on examination of the patient. An extreme mechanistic approach is overemphasized by the constant reference to the human body as a 'physical machine' which must be got into a 'good shape of working efficiency'. Modern physical methods of treatment such as ECT and insulin coma are dubbed under the global title of 'The Drastic Therapies'. Among these the prominence and favour given to transorbital lobotomy using an electrically induced convulsion as an anaesthetic would hardly be viewed with favour in this country.

In spite of these criticisms this book contains much sound common sense. The medicolegal aspects of American psychiatry are well and interestingly dealt with. Psychiatrists who read it will probably be less likely to overlook coexistent physical disease in their cases. The stress on pharmaceutical therapy of mental disorder reminds us that drugs can control symptoms while other methods are given time to take effect. The book rightly emphasizes the role of the family doctor in the treatment of mental disease; and it is he, rather than the psychiatric specialist, who may find the message in these pages of value.

J. A. HARRINGTON

The Psychological Variables in Human Cancer. Edited by J. A. GENDERELLI and F. J. KIRKNER. California University Press and Cambridge University Press. 1955. Pp. iv + 135. 22s. 6d. net.

This book reports a symposium held in 1953 in California, on investigations which tried to evaluate the relationship, if any, between human personality variables and neoplasms. That such a symposium should have been held at all will surprise many. The editors recognize that this is a difficult subject prone to misinterpretation and error, and rightly stress that the data presented should be viewed in a critical light until further research has been carried out. The main discussion centres round observations alleged to show that differences in the host's resistance to the growth and spread of malignancy are correlated with patterns of behaviour. The hypothesis is that the rapidity of the disease process, as measured primarily by survival time, correlates significantly with anxiety, defensiveness, and inhibition, as measured by psychological tests, chiefly the MMPI. Predictions using MMPI profiles were shown to be 78% correct in assessing fast- and slow-growing neoplasms. The classification into slow- and fast-growing tumours leaves much to be desired, and there are obvious criticisms centred round the validity of the psychological tests employed. Even assuming the MMPI to be a valid test for the purpose, and the correctness of the experimental findings, there are obviously other ways in which the results can be interpreted. The experiments deal with so many uncontrolled variables that spurious correlations may readily arise; observations on relationships do not establish causality. Even if it is true that a state of anxiety coincides with an increased malignant growth activity and a state of serenity with its retardation, it remains undecided which factor came first. Many of these criticisms are contained in the discussions following the papers.

Among the papers quoted, that by Schefflin on the incidence of cancer in institutionalized psychotics is of

interest. He found the incidence of cancer in schizophrenics about half that in the general population and occurring at a later age; an exception being the paranoid schizophrenic in whom cancer occurs about four times as frequently as in the general population. The implication that cancer growth may be influenced directly or indirectly by psychological factors should not be taken too seriously in the present stage of our knowledge. Theories to account for relative immunity are legion and there are good *a priori* reasons to suppose that the solution will not come from the field of psychosomatic medicine.

The book makes one thing clear: that we lack the proper tools and techniques to investigate such a problem adequately at present. Until we do, it is doubtful if further extensive research in this field is justified.

J. A. HARRINGTON

Psychomotor Aspects of Mental Disease. An Experimental Study. By H. E. KING. Cambridge, Mass.: Harvard University Press; London: Geoffrey Cumberlege. 1954. Pp. xiv + 185. 28s.

This is an interesting experimental study of psychomotor ability in chronic mental patients. A battery of objective tests was used, including reaction time, speed of tapping, and finger dexterity, and some 300 patients and controls were tested in all. Very marked differences were found, and the author quite rightly feels that the defects in mentally disordered patients unearthed here are of fundamental importance, from both the theoretical and the practical point of view. His conclusions certainly fit in well with those of others who have used objective behaviour tests on neurotic and psychotic groups of patients.

It is doubtful if this study really is worthy of presentation in book form at the price of 28s. The text is padded, partly by lengthy discussion, partly by excessive documentation in the form of figures and diagrams. The author justifies this on the grounds that it makes comprehension of numerical material easier for the uninitiated. This seems a doubtful argument; anyone unfamiliar with means and standard deviations is unlikely to derive much benefit from diagrammatic representations either.

H. J. EYSENCK

Psychotherapy and Personality Change. By CARL R. ROGERS and ROSALIND F. DYMOND. Chicago: University of Chicago Press. 1954. Pp. ix + 447. 45s.

The publication of this book, which reports some results of the Chicago experiment on the effects of psychotherapy, has been awaited with impatience by many psychologists who had heard of this ambitious project, or who had seen preliminary communications. Its final appearance will provide psychologists and psychiatrists with considerable material for thought and for discussion. In essence, what the authors have done has been to select a small group of clients who came to the University of Chicago Counselling Centre, administer a variety of tests to the subjects before and after therapy, and then study the changes in test performance which had taken place. Main reliance was placed on the Thematic Apperception Test and on personality questionnaires, the latter being presented in the somewhat novel form called *Q-sort*, in which the subject, instead of answering 'Yes' or 'No' to a number of statements regarding his personality, has to sort these statements in order of applicability to himself. No objective behaviour tests were administered.

An attempt was made to procure control groups along two lines. One was to have a group of normal people who were also given the tests, but, of course, who received no therapy; the other method used was called the 'own-control' group, in which patients are tested first, then kept waiting for 60 days without therapy, are tested again, then receive therapy, and are tested once more. A last test is applied to all these groups after a follow-up period of 6-12 months.

The outcome of the experiment at first sight seems to be very positive. There are considerable changes for the better after therapy; no such changes are reported for the control group; and only very small changes are found in the 'own-control' group during the waiting period. These results, however, cannot be accepted at face value. In the first place, the measures taken, i.e. questionnaires and T.A.T., are highly subjective and presumably subject to suggestion and social pressure; they would be acceptable as part of a larger battery, including objective and physiological measures, but they certainly are not adequate by themselves to provide acceptable proof of the effects of psychotherapy. Secondly, the authors fail to provide a proper control against which the improvement of the therapy group could be measured. The treatment of this whole problem is disingenuous. A control group made up of normal people obviously will not do; there is no reason to expect from them any improvement from the normal status simply because of the passage of time. As regards the 'own-control' groups, this might at first sight seem to present an adequate

control because the same patient's progress is assessed during a non-treatment period and during a treatment period. This would be true only, however, if the two periods were comparable in length. I have been unable to find any detailed statement regarding the length of treatment period, but the impression given is that this period was considerably longer than the 60 days of non-treatment. Consequently, we might expect the effect of spontaneous remission to be much greater during the lengthy treatment period than during the short non-treatment period, so that the effects of treatment and spontaneous remission would become confounded. This would appear to be the greatest weakness of this experiment and one which makes it impossible to accept the results at face value.

A third argument is likely to be put forward by psychiatrists more familiar with the more serious types of neurotic disorder. Most of the 'clients' taking part in the experiment appear to have been suffering from very minor forms of disorder, which in this country, at least, would not be dignified by the term 'neurosis', and which would not, by and large, receive or request any kind of psychotherapy. One cannot necessarily argue from effects produced in what appear to have been fairly normal people, largely under situational stress, to what might happen with seriously ill neurotic and psychotic patients. The reviewer finds it difficult to evaluate this third point as no detailed evidence is given by the authors of this book on this crucial point.

In discussing what appear to be the three main criticisms of the experimental design, no mention has been made of a large number of small criticisms, most of which are mentioned by the authors themselves, and which may have a very powerful cumulative effect. There is no room to discuss these in detail, but the reader will be aware that at each point experimental rigour has been sacrificed to the putative interests of the clients. The decision to do so was based on ethical grounds which it would be pointless to discuss.

Carl Rogers, in his introduction, provides a fitting comment on this work. Adapting a saying by Robert Hutchins he says: 'It isn't a good research in psychotherapy; it's just the best that there is.' The reviewer cannot disagree with the first part of this sentence, but the fact that he is forced to agree with the second part is an adequate comment on the work that has hitherto been done in this important area of psychological research.

H. J. EYSENCK

Group Therapy for Mothers of Disturbed Children. By HELEN E. DURKIN, Ph.D. Oxford: Blackwell Scientific Publications Ltd. 1954. Pp. xiv + 125. 25s.

This short book by a lay therapist, who is also a lecturer in psychology, reports about 15 years of experience of a psychoanalytically oriented method of therapy applied in about 100 groups of mothers meeting once a week. In the author's experience this therapy contributes a great deal to the improvement of relations within the family, thereby helping and safeguarding independent therapeutic work with children. Any serious effort towards these ends must claim attention from child guidance clinicians and family case workers.

The work reported is clearly that of a sensitive and experienced therapist, and the discussion and brief case illustrations contain many insights and suggestions which group therapists will find useful. Of particular interest are some suggested criteria for selecting members of these groups, the use made of transference relations within the groups and the discussion of the suitability of this form of therapy for 'character analysis'.

But taken as a whole the book falls far short of the claims made for it in Slavson's brief and somewhat pontifical Foreword. It makes only very limited contributions towards the development of a rationale for group psychotherapy, and the possible differences between the approach used with the mothers' groups and that in group therapy in general are not discussed at any length. We would like to understand much more about the effect on the therapeutic process of orienting the group work towards the mother-child relation, for the use of this fundamental relationship as a focus of interpretation, in transference and outside the group, may have important implications for therapeutic practice.

H. PHILLIPSON

Group Discussion. Edited by JOHN BURTON. London: Central Council for Health Education. Pp. 92. 3s. 6d.

The theme of this short symposium is that of the merits of education by discussion as against the passive nature of more formal didactic educational methods.

It opens with several articles on the organization and running of discussion groups in general, and follows these up with an account of varied applications of the group discussion technique. The advice in the

opening articles is presumably addressed to the would-be group leader and ranges from the formal procedure involved in running group meetings to thumbnail sketches of types of group members with which the leader might have to deal (e.g. The wanderer, the silent, the shy, the monopolizing, etc.). The remainder of the symposium deals with the application of group discussion to the teaching of school children, youths, students, the general public and industrial workers, and closes with a discussion of group psychotherapy.

The majority of these articles, although interesting and informative, overemphasize the role of the leader in such a way that he is too often seen as an all-seeing, omnipotent eye functioning outwith the group. The point that the leader is at the same time a member of the group, subject to the same forces as other members, tends to be neglected. An article by Dr Polanie on the teaching of students by group discussion does deal more cogently with the function of the leader. He warns against the danger of the leader viewing the discussion as a group tutorial and stresses the leader's main aim as being—'to merge with the group and disappear'. Dr Anthony's discussion of group analysis contains much that is relevant to the functioning of any group of people but, unfortunately, this is the last article in the symposium and its relevance to other types of activity groups previously mentioned may not be too obvious.

Two articles merit particular mention for their originality and interest. Miss Kelbie describes an interesting technique of spontaneous group conversation used first with post-war refugees in Germany. Here the members of the group are given the general theme of childhood memories at a particular age and encouraged to make personal contributions. She points out the advantage that all can participate easily and that the happy memories shared are easily projected into the group to promote positive friendly group feelings. The dangers of negative feelings resulting are also discussed. The basis of this technique has much to link it with that of group analysis in a more superficial and direct form.

After the initial articles dealing with the form of organization and running of discussion groups and committees, etc., and detailing the different roles of leaders and members within these groups, it is somewhat of a relief to read Beatrice Pollard's interesting account of the Quaker Group: 'There are no rules, no standing orders, no rules of procedure, no motions, no commitments, no closure motions or any of the other apparatus with which the world of committees is familiar.' The Quaker group, in fact, appears to represent a halfway-house between the formal business meeting or committee type of group and the purely psychoanalytic group.

A. MCGHIE

Time Distortion in Hypnosis. By LINN F. COOPER and MILTON H. ERICKSON. Baltimore: Williams and Wilkins Co.; London: Ballière, Tindall and Cox. 1954. Pp. ix + 191. 31s. 6d.

Part I (158 pp.), by Cooper, reports 'pilot experiments' (some previously published) with fourteen subjects, mostly students, on phenomena during 'time distortion' and 'time manipulation' in hypnosis: viz. injection of sound signals into hallucinatory activities; learning, motor and non-motor; review and practice of previously learned material; association; mathematical and creative mental activity. 'Time distortion' is defined as increase of seeming duration over clock reading, although the subject experiences the normal rate of activity; its most important implication is the possibility of *accelerated mental activity*, which would obviously be of tremendous practical significance.

Part II (28 pp.), by Erickson, describes therapeutic application in 6 patients.

This is at best an amateurish attempt at research. There is no reference to the experimental work on time perception, hypnosis, thinking, etc., only to the authors' own, 'communications', 'suggestions', etc. No objective evidence of accelerated mental activity is adduced, only anecdotal data, e.g. counting 821 jelly beans from a shoe box in a seeming duration of 10 minutes, actual time 10 seconds. The clinician's defence is the false dichotomy between scientific and human interests. The fundamental fallacy is that time is regarded as a *thing* to be measured instead of as itself a *measure*, 'experiential' time being but an individual's estimate of 'physical' time. Since its nature is not yet understood, to use hypnosis as instrument instead of object of research is premature. What these workers have done is merely to have tested the effectiveness of their hypnotic suggestion, contributing nothing towards understanding the status of hypnotic experiences.

J. E. MACDONALD

The Hands of Children: an Introduction to Psycho-chirology. By JULIUS SPIER. Second Edition, with an Introduction by C. G. JUNG. Translated by VICTOR GROVE. Illustrated with 86 drawings by F. W. W. SIMON and 69 Plates. London: Routledge and Kegan Paul Ltd. 1955. Pp. xvi + 199. 28s. net.

The first edition of this work was published in 1944. The new edition contains a second appendix by Herta Levi on the hands of the mentally diseased, with 46 plates to illustrate it, in addition to the first appendix on specimen analyses on the hands of children, which is illustrated by 23 plates.

The body of the book consists of eleven chapters setting forth the theoretical foundations of psycho-chirology and its application to the hands of children. To the empirical and experimental psychologist a book of this kind offers many problems and little help. It has developed out of traditional chiromancy and uses its language. Like graphology, it has adopted Jung's psychology as its theoretical background, and claims to be the intuitively established basis of a scientific approach to the psychology of personality. Exactly how much of it would survive strictly scientific and experimental tests, if they could be applied, remains problematic, though it contains numerous significant observations. Herta Levi shrewdly points out in the preface to the second edition that, as a scientific method, chiromancy is still in its infancy and requires a good deal of body-building nourishment in the shape of factual data and statistics, but adds that very few of us have the great intuitive powers of a Julius Spier, and therefore have to rely more on scientific methods and exchange of information!

R. W. PICKFORD

Low Intelligence and Delinquency. By MARY WOODWARD. London: Institute for the Study and Treatment of Delinquency. 1955. Pp. 24. 1s. 6d.

This pamphlet reviews a 40-year old controversy on the relation between low intelligence and delinquency. There is also a brief report of an intensive study of a small group of delinquents. The conclusion is that 'low intelligence cannot be regarded as an important causal factor in delinquency': whether intelligence has any relation at all to delinquency is not settled, owing to the difficulty of controlling the cultural factors which accompany delinquency and depress test scores. This conclusion is probably not surprising; but the pamphlet performs a valuable service in its concise review of the literature, summarizing 66 references in 16 pages. On this topic, successive investigators obtained conflicting results, a state of affairs not unknown to the social sciences, which tends to discredit research. The writer tries to resolve inconsistencies and present findings from the accumulation of results.

J. D. NISBET

Culture and Human Fertility. By FRANK LORIMER. UNESCO. 1954. Pp. 514. 25s.

The book divides into two sections, the first being a theoretical inspection of the relation of cultural aspects and human fertility, while the second presents a number of special studies in this field contributed by other collaborators. The reports in the latter section illustrate clearly the pitfalls which beset the unwary purveyor of vital statistics in areas where not only are reliable figures unavailable but where the population may view with darkest suspicion any attempt at enumeration. Prof. Mortara points out this absence of reliable statistics for Brazil, and the other studies reveal the perplexities arising for workers in preliterate communities in this respect. Further, the question of fertility has great emotional significance to many peoples, and this may lead to distortion or suppression—if not complete misrepresentation—of facts. In short, local conditions and attitudes may combine to produce data of low validity. Thus the need for caution in interpreting the results of such studies and the need for other carefully regulated investigations to supplement them must be emphasized.

In the theoretical section Prof. Lorimer draws on a wide range of material. But, as in much comparative work at the present stage, his exposition tends to show up the gaps and uncertainties in our knowledge. Believing in the pervasiveness of sociological factors, the reviewer agrees with his main thesis that these factors may exercise a profound influence on human fertility. The studies quoted appear to indicate this assumption. However, at present it seems impossible to do more than to formulate fairly general hypotheses, as he does.

The book, then, serves as a valuable exploratory work. The evidence is at times a little thin and the statistical material somewhat dubious. It must be considered as blazing a trail, not as a definitive assertion of the complex inter-relationships between culture and human fertility.

MARY BIRD

Sex and Society. By KENNETH WALKER and PETER FLETCHER. Harmondsworth: Penguin Books. 1955. Pp. 251. 2s. 6d.

This book is concerned largely with the question, 'What is wrong with sex in the modern world?' The authors take the view that much is wrong, and that sexual disorders are symptoms or aspects of unsatisfactory conduct of life in general, which in turn is due largely to the inhibiting and distorting effect of attitudes widely engendered in a highly organized 'mass society' upon individual spontaneity and values. This is regarded very seriously indeed—as 'a matter of life or death' for our civilization. While one may have sympathy for the general thesis (which is by no means new) one may well doubt whether the situation is as bad as the authors make out, and whether the outcome is likely to be as disastrous.

In the course of the book, many sensible things are said, as in the chapter 'Sexual Morality before Marriage', and some less sensible things, as in the chapter 'Introduction to Neurosis'. There is a tendency for 'typical' cases to be referred to without any evidence as to how typical they are, and for statements to be made without adequate basis being indicated; for example, 'It is not surprising that emotional maturity is a rare phenomenon in the modern world.'

D. GRAHAM

Working-class Anti-Semite. By JAMES H. ROBB. London: Tavistock Publications Ltd. 1954. Pp. xiv + 239. 15s. net.

In this study the author interviewed a near-random sample of 103 Bethnal Green working men and rated them on anti-Semitism and a number of personality variables, such as degree of social isolation. He obtained information on experience of unemployment, presence of Jewish workmates and so on; and attempted to relate the anti-Semitic personality to the background of the subculture of the district.

The results show that only one man in eight of the population is completely tolerant, but over half the remainder have to be openly invited to express their anti-Semitism before they will do so. The correlations between anti-Semitism and the personality variables rated were modest (0.33 to 0.57) but in the expected direction; no significant relations were found between anti-Semitism and any of the environmental factors investigated. In general the findings, such as they are, do not dispute any of the current views in this field.

The most interesting and potentially useful technique to come from this work, is the author's method of rating anti-Semitism according to the amount of provocation required to elicit a prejudiced remark. This variant of the focused interview would seem to have many uses, not in this field alone. It is thus all the more unfortunate that no attempt was made to validate the *a priori* classification used against, say, questionnaires or systematic case studies. The problem of the validity of making personality assessments from a 25-minute doorstep interview is nowhere tackled; and most of the correlations presented are suspect on the grounds of non-independence of the raters.

One further inadequacy must be noted; the author's attempt to relate the psychological and the sociological material. This is done by means of the concept of 'typical personality'. For the social psychologist, whose interest is in measurement in the social sciences, 'typical', if it means anything, should mean the mode of a modal distribution. Its use should be restricted to cases where this distribution has been first established; otherwise explanation in these terms becomes too easy. 'So easy', as Cattell said in another connexion, 'as to be worthless'.

In spite of these criticisms one must welcome this book as the first attempt to study anti-Semitism in this country on anything other than the questionnaire level. In a field where we have happily become used to the sort of on-going team researches of the California and Maudsley groups, it is inevitable that the contribution of one man working for two years should be considered as little more than a brave try. The book does serve, however, to emphasize the need for co-operative research and highlights the difficulties of a psychologist tackling this task alone.

DEREK PUGH

Christian Life and the Unconscious. By Dr ERNEST WHITE. London: Hodder and Stoughton. 1955. Pp. 190. 10s. 6d.

In assessing the value of any book two things need to be kept in mind: (1) What it contains—not what it doesn't; (2) For whom it is written. Dr White's book is an attempt at showing how the Christian faith can influence the levels of the unconscious through the operations of the Holy Spirit. He distinguishes

between Conversion and Re-Birth which works 'secretly' in the true Convert's mind. On the whole the book is orthodox on the religious side, and shows a distinctly eclectic analytical approach on the psychological. But the author has in mind Christians who are puzzled by the apparent dichotomy between religion and modern psychology. Readers who are biased against the former will find it difficult to appreciate the fundamental premises of the author, and those prejudiced against the latter are likely to reject the necessity for defining the mind in the way he does. This does not mean that the work is without merit, because it has considerable value for those who are willing to read it sympathetically. At times one gets the idea that Dr White's own religious beliefs intrude upon his work as a medical psychologist, but it is hardly likely that he is guilty of proselytizing, for he must know that no lasting good comes from the imposition of ideas on sick minds. The whole aim and purpose of the book is positive in the sense that it shows how religion can be aided by psychology, and how the human personality can be enriched through faith and religious experience.

W. E. SARGENT

A Dictionary of Pastoral Psychology. By VERGILIUS FERM. New York: The Philosophical Library. 1955. Pp. xi + 336. \$6.00.

The comprehensiveness of this book and its breadth of description could earn for it the perhaps more impressive title of encyclopaedia. In addition to the large number of psychological terms adequately defined, many subjects within the field of pastoral psychology are expansively treated, e.g. Pastoral counselling occupies 41 pages—a small book on its own. The volume is fairly well up to date, and could be extremely helpful to ministers in search of reliable principles on which to work when confronted with human problems, or in understanding new concepts relative to a pastor's life. The many cross references lend clarity and there are pertinent suggestions for further reading. A number of authorities, all specialists in their own fields, have contributed specific articles. Being an American, the author naturally quotes many of his own countrymen; one wishes, however, that a little more room had been given to writers in the British field of Pastoral Psychology.

W. E. SARGENT

Answer to Job. By C. G. JUNG. Translated by R. F. C. HULL. London: Routledge and Kegan Paul. 1954. Pp. xviii + 194. 12s. 6d. net.

This book makes difficult, but rewarding, reading to the reader who has some acquaintance with Jungian thought. To the reader who is not familiar with that discipline, it appears strange and puzzling. To any reader, it would be much more rewarding if the translation were better: the language is uneven, ranging from the obscure to the colloquial, and it often jars unnecessarily.

The theme centres round the existence of evil, and the development of God's nature and man's, both of whom need awareness of their own darker side in order to achieve wholeness. Job, by his greater awareness and insight, offers a challenge to God, and the book thereafter, with great erudition and width of allusion, traces the development of self-awareness and of the archetypes in God—good and evil, the masculine ideal of perfection and the feminine ideal of wholeness—through the prophecies, the Incarnation, the Apocalypse and finally the doctrine of the Assumption of the Blessed Virgin Mary. The psychological aspect of the latter is interesting and important and dealt with here for the first time.

Atomic Warfare, and the way in which man's intellectual development has outstripped his spiritual growth, are shown as evidence of the urgent need to consider these matters.

Jung himself points out that the psychology of religion has two aspects, the psychology of religious persons and the psychology of religious 'contents'. He has himself, in this book, made a rare and original contribution to the latter.

A. M. SILVER

Olive Schreiner: a Study in Latent Meanings. By MARION V. FRIEDMANN. Johannesburg: Witwatersrand University Press. 1954. Pp. vii + 67. 8s. 6d.

This book is an interesting attempt to apply psychoanalytic hypotheses to the interrelationships of neurosis and creative art. The personality of Olive Schreiner, as partially reflected in the characters and action of her novels, and as shown in her letters and life history, is dissected with some skill. These two sources of evidence are functionally correlated in the chapter on 'Latent Meanings'.

The study suffers from the usual disadvantages inherent in its approach. Any unusual work—whether ‘art’ or not—is generally the product of an unusual personality—statistically ‘abnormal’. It is in the application of qualitative criteria of abnormality, therefore, that the approach adopted by Miss Friedmann may be valid. The neurotic is, compared to the artist, relatively helpless in the face of his difficulties, and resorts to inadequate forms of behaviour, which the clinician has little hesitation in labelling as pathological. There is some doubt in the mind of this reviewer whether this word can, without undue simplification, profitably be applied to the creative activity of the successful artist.

In the present instance, Miss Friedmann has thrown considerable light on the reflection of Olive Schreiner’s own life in her fiction, though little is made of the political works and influence of the woman who has been called ‘the first suffragette’. Reaction against oppression, even though unusual in the South Africa of 1900, cannot be regarded simply as maladaptive behaviour.

This book does not nor can it be expected to, answer the question ‘Why was Olive Schreiner a great artist?’, but it illuminates what she wrote, and contributes to our knowledge of what her writing meant to her. This year is the centenary of her birth and this study, though slight, is a not entirely unworthy monument to its subject.

S. G. LEE

The Function and Training of Mental Nurses. By A. N. OPPENHEIM. London: Chapman and Hall. 1955. Pp. 85. 12s. 6d. net.

Twenty years ago the duties of mental nurses were restricted to the custodial care of the mentally sick. With the introduction of complicated physical treatments, the mental nurse has had to fulfil duties more like his or her general trained counterpart. More recently, mental nurses in some Units have had to fulfil the duties of ‘social therapists’, in which the traditional skills of the hospital nurse are no longer relevant. It is little wonder then, that some confusion prevails as to the functions of the mental nurse. This in turn has an urgent bearing on training.

Mr Oppenheim’s occupational study of mental nursing at the Bethlem Royal and Maudsley Hospitals is therefore timely, and it is not surprising that it has received considerable publicity in the Nursing periodicals. His study has been thorough and painstaking, and could well be taken as a model for similar studies in other hospitals.

According to an Appendix on Formal Training at the end of the book, the 16 lectures on Psychology included in the nurse’s three-year course are given by the Sister Tutor and by ‘a doctor’. One wonders why Clinical Psychologists are not asked to give any of these lectures.

RALPH HETHERINGTON

